Regular configurations and TBR graphs

Wu, Taoyang

Regular Configurations and TBR Graphs

Taoyang Wu
Queen Mary, University of London

Thesis submitted to the University of London
for the degree of Doctor of Philosophy

May 2009

Abstract

This thesis consists of two parts: The first one is concerned with the theory and applications of regular configurations; the second one is devoted to TBR graphs.

In the first part, a new approach is proposed to study regular configurations, an extremal arrangement of necklaces formed by a given number of red beads and black beads. We first show that this concept is closely related to several other concepts studied in the literature, such as balanced words, maximally even sets, and the ground states in the Kawasaki-Ising model. Then we apply regular configurations to solve the (vertex) cycle packing problem for shift digraphs, a family of Cayley digraphs.

TBR is one of widely used tree rearrangement operationes, and plays an important role in heuristic algorithms for phylogenetic tree reconstruction. In the second part of this thesis we study various properties of TBR graphs, a family of graphs associated with the TBR operation. To investigate the degree distribution of the TBR graphs, we also study Γ-index, a concept introduced to measure the shape of trees. As an interesting by-product, we obtain a structural characterization of good trees, a well-known family of trees that generalizes the complete binary trees.

Contents

List of Figures

Part I

Regular Configurations

Chapter 1

Introduction

As the title suggests, the purpose of Part I in this thesis is to explore the theory and applications of regular configurations, an extremal class of configurations on cycles. We start with an informal description while the formal definition will be presented in Chapter 2.

Given a necklace with d beads, how many ways can we color it with two colors, say red and black, such that there are exactly a red beads and $d-a$ black beads? As a classical counting problem in combinatorics, the reader can work it out in a few minutes, or find the answer in Section 1.4.2 as well as in many combinatorics books [59].

Informally speaking, a configuration in the Kawasaki-Ising model is such a coloring. Instead of the above counting problem, here we are interested in a family of extremal colorings: the ones such that the distribution of the two colors on beads is as evenly as possible.

This family of colorings, called regular configurations, is one of the main objects in this part. They are important for many theoretical studies and practical applications. For example, if we regard the arrangement of white keys and black keys on a piano keyboard as a configuration, then it is a regular one (see Section 1.4.3 for more details).

The above example is one motivation for musicians to study regular configurations, where they are called maximally even sets [16]. Independently, they are also rediscovered in symbolic dynamics (combinatorics on words), where they are referred to as (cyclic) balanced words, a finite version of Sturmian words [37].

Informally, this part shows how regular configurations, a natural combinatorial object, could be discovered from studying a problem in graph theory. This problem, called cycle packing, is to calculate the maximum number of vertex-disjoint cycles in a given digraph. It is an important problem, and is closely linked with guessing number, one parameter of graphs studied in this part.

Though they have been extensively investigated in many different fields, regular configurations are studied here from a new perspective: characteristic sequences. For each configuration, this sequence provides a parameter to measure its regularity, and regular configurations are the ones with maximal regularity.

There are many "nice" properties associated with regular configurations. For example, they are unique, in the sense that there exists one and only one regular configuration up to rotation for given a and d. They are self-similar, that is, a configuration is regular if and only its "characteristic configuration" is regular. These properties, as well as many others, are investigated here from this new perspective.

Despite evenness, there are several other reasons to call regular configurations extremal. One of such reasons can be found in the works of Jenkinson and his coauthors [37, 34, 36, 35]. Furthermore, they can be characterized as ground states, the configurations with the minimal energy in the Kawasaki-Ising model [23]. Two new proofs of this fact are given in this part: one shows the connections between regularity and the Hamiltonian; the other shows one dynamic aspect of the Kawasaki-Ising model, i.e., how non-regular configurations can been evolved to the regular ones.

1.1 Background

In this section, we give a brief discussion on the background of some notation mentioned so far. The first one is for configurations. Here we mainly follow Biggs's book [7].

Configurations

In fair broad generality, configurations can be defined as follows.

Definition 1.1.1. *Let Λ be a finite set, and G be a graph. A configuration σ is a map from the vertex set $V(G)$ to the set Λ defined as*

$$\begin{aligned} \sigma : \qquad V(G) &\longrightarrow \Lambda \\ v &\longrightarrow \sigma_v. \end{aligned}$$

Here σ_v, also denoted by $\sigma(v)$, is an element of Λ. The set of all such configurations will be denoted by $\Omega(G, \Lambda)$, or just Ω when G and Λ are clear from the context.

The generality of this definition comes from two aspects: the interpretation of the finite set Λ and the restriction on the map. For instance, if Λ is a set of colors, then the map σ is a vertex coloring, i.e., an assignment of colors to the vertices of G. If we further require that $\sigma_u \neq \sigma_v$ for any edge (u, v) in G, then σ is a proper vertex coloring.

Sometimes it is convenient to endow the set Λ with some algebraic structure. One such structure of special interest is that of a 'ring'. That is, we allow two operations, $+$ and $\times$, on the elements of Λ. In this case, the configurations σ in $\Omega(G, \Lambda)$ may themselves be combined by operations derived from the structure of Λ. In fact, let σ and ϕ be two such configurations. Then we have

$$(\sigma + \phi)(v) = \sigma(v) + \phi(v) \quad \text{and} \quad (\sigma \times \phi)(v) = \sigma(v) \times \phi(v).$$

Thus the set Ω becomes a ring itself. In particular, if Λ contains 0 and 1, then there are two configurations, $\mathbf{0}$ and $\mathbf{1}$, such that $\mathbf{0}(v) = 0$ and $\mathbf{1}(v) = 1$ for each v in $V(G)$. And it is straightforward to verify that **0** and **1** are respectively the identity element for $+$ and the identity element for $\times$ in Ω.

Here we want to stress that configurations have slightly different meanings in different parts of this part. Roughly speaking, in Section 4.4, $\Lambda = \mathbb{Z}_2$ and the algebraic structure plays a crucial role. In other place, $\Lambda = \{+1, -1\}$ and the algebraic structure is less important except in Chapter 3. A more detailed study of configurations is given in Section 1.4 where we will focus on the Kawasaki-Ising model, a variant of the well-known Ising model.

Balanced words

A word is a (possibly infinite) sequence of symbols drawn from a finite alphabet, say $\{0, 1\}$. Any finite contiguous subsequence of a word w is called a *factor*.

The set of factors is denoted by $L(w)$ and the set of factors of length $n \geq 0$ is denoted by $L_n(w)$. First studied by Morse and Hedlund [45] in symbolic dynamics, Sturmian words are aperiodic infinite words over $\{0, 1\}$ that are balanced: Denoting the number of occurrences of i ($i \in \{0, 1\}$) in w by $|w|_i$, then a word w is called *balanced* if we have $|u| = |v| \Rightarrow ||u|_0 - |v|_0| \leq 1$ for all $u, v \in L(w)$.

In this thesis, we are mainly interested in finite cyclic words, i.e., their first letter and last letter are considered to be adjacent. Therefore all words mentioned later are cyclic unless otherwise stated.

The standard reference for Sturmian words is Berstel and Séébold's chapter in Lothaire's book [43], which also contains a historical account. For finite words, one recent survey has been conducted by Berstel and Karhumäki [6].

Here we propose to study words from the perspective of configurations. More precisely, the alphabet set will be regarded as the spin set Λ. Therefore, an infinite word is a configuration on the one-dimensional lattice, while a finite word of length k is a configuration on the cycle C^k. More details about this connection will be explored in Section 2.2 and

Chapter 3.

Shift Digraphs

A *shift digraph*, $\Gamma = \mathrm{Cay}(n; \{\alpha, \beta\})$, is a Cayley digraph of $\mathbb{Z}_n$ with two generators, say α and β. To avoid degenerated cases, we will assume $0 < \alpha < \beta \leq n-1$ throughout this part. More precisely, the vertex set of Γ is $V(\Gamma) = \{0, 1, \cdots, n-1\}$ and the arc set is $A(\Gamma) = A_\alpha \bigsqcup A_\beta$, where $A_\alpha = \{(i, i+\alpha) \pmod n \mid 0 \leq i < n\}$ and $A_\beta = \{(i, i+\beta) \pmod n \mid 0 \leq i < n\}$. The arcs in A_α are called type I while the arcs in A_β are called type II. They are generated respectively by α and β.

One case of special interest is when $\alpha = 1$ and $\alpha < \beta < n-1$. In this case, $\mathrm{Cay}(n; \{1, \beta\})$ is also written as $\mathrm{Shift}(n, \beta)$, and is called a *directed double loop*. Their underlying graphs, the Cayley graphs of $\mathbb{Z}_n$ with two generators 1 and β, are also referred to as double loops, cyclic graphs or chordal rings in the literature [26, 41, 5, 44], and have a vast number of applications to telecommunication network, VLSI design and distributed computations [5, 13, 40, 44].

More generally, the underlying graphs of shift digraphs, the Cayley digraph $\mathbb{Z}_n$ with two generators $\{\alpha, \beta\}$, are a special family of circulant graphs. Such graphs have been intensively studied [10, 18, 24, 42, 67].

Note that the cycles in shift digraphs can be coded as configurations. More precisely, a cycle with length d can be coded as a necklace with d beads, or a configuration on C^d. As we will see later, this connection is our first step to study the cycle packing problem by regular configurations.

1.2 Outline

Part I of this thesis investigates regular configurations, both its theory and applications. Besides the current chapter on background and definitions, it consists of the following three chapters.

Chapter 2 is devoted to regular configurations, the main object in this part. After introducing the degree of regularity, we present the formal definition of regular configurations. The remainder of this chapter is about a variety of properties of regular configurations, including self-similarity and symmetry, and the connections to balanced words and maximally even sets. Some results in this part are contained in the work of the author [64].

Chapter 3 presents another characterization of regular configurations by ground states in the Kawasaki-Ising model. This is obtained by two different approaches. One shows the links between regularity and energy. The other approach deals with the dynamics aspect of the Kawasaki-Ising model. As a byproduct, we also obtain stability lemmas, which give us certain structure information about semi-regular configurations. Some results in this part have appeared as a joint work of the author and Peter Cameron in [14], a preliminary version of [15].

Chapter 4, is about the cycle packing problem in shift digraphs. By regular configurations, we prove that the cycle packing number of a shift digraph D depends only on its size and girth. This result is also used to study the guessing number of shift digraphs. Some results in this part are contained in the author's work [64, 65] and the results concerning guessing number will appear in [58], a joint work with Peter Cameron and Soren Riis.

1.3 Notation and terminology

In this section, we will fix some notation and terminology that will be used throughout this part. Here we are mainly following [20] and [4].

The set of positive integers, nonnegative integers, integers and real numbers will be denoted by $\mathbb{N}^+$, $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$ respectively.

Given $a, b \in \mathbb{N}$, we say a divides b and denote it by $a|b$ if $ax = b$ for some $x \in \mathbb{N}$. In this case, we also say a is a *divisor* of b. The greatest common divisor of a, b, denoted by $\gcd(a, b)$, is the greatest number t such that $t|a$ and $t|b$. Similarly we can define $\gcd(a, b, c)$ for a triple of positive numbers.

For $n \in \mathbb{Z}$, let $[n]$ denote the finite set $\{0, 1, \cdots, n-1\}$. We also associate it with an ordering $<$ in $[n]$ as follows: $0 < 1 < \cdots < n-1$. Given $x \in \mathbb{N}$, let $(x)_n$ denote the integer in $[n]$ which congruent to x modulo n. For $x \in \mathbb{R}$, let $\lfloor x \rfloor$ denote the greatest integer $\leq x$ and $\lceil x \rceil$ denote the least integer $\geq x$. Furthermore, let $\{x\}$ denote $x - \lfloor x \rfloor$.

The set of $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n is denoted by $\mathbb{Z}_n$. For brevity, the elements in $\mathbb{Z}_n$ will be written as $\{0, 1, \cdots, n-1\}$ as well, which should cause no confusion with $[n]$ from the context. Alternatively, $\mathbb{Z}_n$ is the cyclic group of order n. The multiplicative group of units of the ring of integers modulo n is denoted by $\mathbb{Z}_n^*$. For later use, we denote by $\mathcal{F}$ one of the elements in $\{\mathbb{Z}, \mathbb{Z}_n, \mathbb{N}, \mathbb{R}\}$.

A finite *sequence* of length n in $\mathbb{F}$ is a map from the set $[n]$ to $\mathcal{F}$. Generally, it is written as $x_0 x_1 \cdots x_{n-1}$ where $x_i \in \mathcal{F}$. A *cyclic sequence* $\mathcal{X}$ is a map from $\mathbb{Z}_n$ to $\mathcal{F}$, and it will also be written as $(x_0, x_1, \cdots, x_{n-1})$. Here we adopt the convention that the subscripts of x_i in $\mathcal{X}$ are calculated modulo n, the *length* of $\mathcal{X}$. For any two elements x_i, x_j in $\mathcal{X}$, their *cyclic distance*, denoted by $d_c(x_i, x_j)$, is defined as $\min\{(i-j)_n, (j-i)_n\}$.

A *shift operator*, denoted by τ, is defined as

$$\tau\mathcal{X} := (x_1, \cdots, x_{n-1}, x_0).$$

This gives us an equivalence relation on cyclic sequences: $\mathcal{X} \sim \mathcal{Y}$ if and only if $\mathcal{X} = \tau^t \mathcal{Y}$ for some $t \in \mathbb{N}$.

To emphasis the importance of the relative positions of the elements in sequences, we will regard a sequence, say $x_0, x_1, \cdots x_{n-1}$, as a *vector* and write it as $\langle x_0, x_1, \cdots, x_{n-1} \rangle$. All vectors of length n form a set, denoted by $\mathcal{F}^n$. If we associate it with an ordering, say $<$, on $\mathcal{F}$, then it will induce a lexicographic ordering, denoted by $<_L$, on the vectors in $\mathcal{F}^n$ as follows: for $X = \langle x_0, x_1, \cdots, x_{n-1} \rangle$ and $Y = \langle y_0, y_1, \cdots, y_{n-1} \rangle$, $X <_L Y$ if there exists an integer $k \in [0, n-1]$ such that $x_k < y_k$ and $x_i = y_i$ holds for $1 \leq i < k$.

A *multiset*, usually denoted by Ξ, is formally defined as a pair (A, m) where A is a set and m maps each element in A to $\mathbb{N}^+$. The set A is called the *underlying set* of Ξ. For $a \in A$, $m(a)$ is called the the *multiciplity* of a in X_i, i.e., the number of occurrences of a in Ξ. Sometimes, we also write a multiset Ξ over a set A as $\{a^{m(a)} \mid a \in A\}$.

Given a cyclic sequence $\mathcal{X} = (x_0, x_1, \cdots, x_{n-1})$ and an element α in $\mathcal{F}$, $\alpha + \mathcal{X}$ and $\alpha\mathcal{X}$ are defined as follows:
$\alpha+\mathcal{X} := (\alpha+x_0, \alpha+x_1, \cdots, \alpha+x_{n-1})$ and $\alpha\mathcal{X} := (\alpha x_0, \alpha x_1, \cdots, \alpha x_{n-1})$. Similarly, these two operations can be defined over sets, vectors and multisets.

A *graph* G is an ordered pair (V, E) consisting of a non-empty set V of *vertices* and a set E of *edges* satisfying $E \subseteq \binom{V}{2}$. Thus, the elements of E are 2-element subsets of V, which are written as uv or (u, v) for some $u, v \in V$. Unless otherwise stated, all graphs mentioned in this part are simple. That is, they have no loops and no parallel edges.

Let $G = (V, E)$ and $G' = (V', E')$ be two graphs. We call G and G' *isomorphic*, and write $G \simeq G'$, if there exists a bijection $\phi : V \to V'$ such that $(x, y) \in E$ if and only if $(\phi(x), \phi(y)) \in E'$ for any pair x, y in V.

We set $G \cup G' := (V \cup V', E \cup E')$ and $G \cap G' := (V \cap V', E \cap E')$. If $G \cap G' = \emptyset$, then G and G' are *disjoint*. In this case, $G \cup G'$ is the *disjoint union* of G and G', and will be denoted by $G \sqcup G'$. If $V' \subseteq V$ and $E' \subseteq E$, then G' is a *subgraph* of G.

A *path* is a non-empty graph $P = (V, E)$ of the form

$$V = \{v_0, v_1, \cdots, v_k\}, \qquad E = \{v_0v_1, v_1v_2, \cdots, v_{k-1}v_k\},$$

where all v_i are distinct. A path is often be presented by the sequence of its vertices. In other words, we will write the above path as $P = v_0v_1 \cdots v_k$ and say that P is a path from v_0 to v_k.

If $P = v_0 \cdots v_{k-1}$ is a path such that $k \geq 3$, then the graph $C := P + v_{k-1}v_0$ is called a *cycle*. To distinguish from paths, we often denote a cycle by its cyclic sequences of vertices; the above cycle C might be written as $(v_0, v_1, \cdots, v_{k-1})$. The *length* of a cycle is the number of edges (or vertices) contained in it; the cycle of length k is called a k-cycle and denoted by C^k.

Given a path P or a cycle C, if P (resp. C) is a subgraph of G, then we say that G contains P (resp. C). The length of the shortest cycle contained in G is called the *girth* of G and denoted by $\omega(G)$.

Given a k-cycle, say $C^k = (0, 1, \cdots, k-1)$, then the interval $[i, i+t]$ for $i, t \in [0, k-1]$ denotes the set $\{i, (i+1)_k, \cdots, (i+t)_k\}$. Given two vertices i, j in C^k, its distance is defined to be $d_c(i, j)$; and $|j-i|_L$ is defined as the minimum positive number s such that $j = (i+s)_k$. Let us remark here that generally we do not have $|i-j|_L = |j-i|_L$.

A *directed graph* (or just *digraph*) D is an ordered pair (V, A) consisting of a non-empty set V of vertices and a set A of arcs, where each of them is an ordered pair of distinct vertices. If $a = (u, v)$ is an element of A, then we say that u is the *tail* of a and v is the *head* of a. The arc a is said to be *directed* from u to v.

In general, the terminology for directed graphs is similar to that of graphs. For example, a *directed path* is a sequence of distinct vertices $v_0v_1 \cdots v_k$ such that there is an arc (v_i, v_{i+1}) for all $i \in [0, k-1]$. A digraph is *acyclic* if it does not contain any directed cycle. Given a digraph D, the maximum number of vertex-disjoint (directed) cycles contained in D, denoted by $\nu_0(D)$, is called the *cycle packing number* of D.

1.4 Configurations

This section is intended to provide a detailed introduction to the Kawasaki-Ising model. We begin with a brief discussion on the Ising model on graphs, a classical model of statistical mechanics. namely the Ising model. For more backgrounds, we refer the reader to Welsh [62].

In the general Ising model on a graph G, each vertex i of G is assigned a *spin*, denoted by σ_i or $\sigma(i)$, which is either $+1$ (called 'up') or -1 (called 'down'). To simplify notation, we also write the up and the down spin respectively as $+$ and $-$. An assignment of spins to all the vertices of G is called a *configuration* or a *state*, and is denoted by σ.

For each edge $e = (u, v)$ of G, we associate it with an *interaction energy* J_1, which is constant. It measures the strength of the interaction between neighboring pairs of vertices. When there is no effect from the external field, the *Hamiltonian* $H_1(\sigma)$ for a state $\sigma = (\sigma_0, \cdots, \sigma_{n-1})$ is defined as

$$H_1(\sigma) := J_1 \sum_{(u,v)} \sigma_u \sigma_v. \tag{1.1}$$

Here we have one ore assumption that J_1 is a positive constant. This means the interactions between adjacent spins are antiferromagnetic.

1.4.1 The Kawasaki-Ising model

In this subsection we will study a variant of the Ising model, the Kawasaki-Ising model. As a fixed-parameter version of the Ising model, it is also called the conserved-order-parameter (COP) Ising model, or Ising gases model, in the literature [48].

The number of vertices in the up spin state in a configuration σ, denoted by $|\sigma|_+$, is called the *weight* of σ. The Kawasaki-Ising model consists of the configurations σ such that $|\sigma|_+ = a$ for a given number $a \in \mathbb{N}$.

In this part we are mainly interested in the Kawasaki-Ising model on the cycle graph C^d. All configurations σ on C^d with weight a form a set, denoted by $\mathrm{KI}(a, d)$. Throughout the part, we will also denote $d - a$ by b. Then (a, b) provides another set of parameters for the Kawasaki-Ising model on C^d. Furthermore, sometimes we also denote $\mathrm{KI}(a, d)$ by $\mathrm{CONF}(a, b)$.

In other words, a configuration σ in $\mathrm{KI}(a, d)$ is a map from $V(C^d)$ to the set of two spins $\{+, -\}$ such that $|\sigma|_+ = |\sigma^{-1}(+)| = a$. Here $V(C^d) =$

$\{0, \cdots, d-1\}$ and the vertices are consecutively labelled. Then any configuration σ can be represented as the cyclic sequence $(\sigma_0, \cdots, \sigma_{d-1})$, which is called the *representing sequence* of σ.

Since σ_i is regarded as an element in a cyclic sequence of length d, the subindex of σ_i is calculated modulo d. In other words, we will write σ_i instead of $\sigma_{(i)_d}$ for $i \notin \{0, 1, \cdots, d-1\}$ when this is clear for the context. Similar conventions will be used for other cyclic sequences.

Note that the ordering $<$ defined in $[d-1]$ induces an ordering on the vertices of C^d, i.e., $0 < 1 < \cdots < d-1$. By this ordering, the vertices in the down spin state of any configuration $\sigma \in \mathrm{KI}(a, d)$ can be enumerated as $\{B_0 < B_1 < \cdots < B_{b-1}\}$, where B_j is the $(j+1)$-th vertex in the down spin.

Denote the number of vertices between B_i and B_{i+1} by x_i. That means

$$x_i = (B_{(i+1)_b} - B_i - 1)_d.$$

Then a configuration σ gives a unique cyclic sequence $\mathcal{X} = (x_0, x_1 \cdots, x_{b-1})$, called the *characteristic sequence* of σ. Let $\mathcal{X}_t = \tau^t \mathcal{X}$ for $t \in [b-1]$. It is clear that each configuration σ is uniquely determined by the pair $(B_i, \mathcal{X}_i)$. Note that if $\mathcal{X} = (x_0, \cdots, x_{b-1})$ is a characteristic sequence for a configuration σ in $\mathrm{CONF}(a, b)$, then

$$x_0 + x_1 + \cdots + x_{b-1} = a. \tag{1.2}$$

For the moment, we will content ourselves with viewing the Kawasaki-Ising model as a combinatorial object, while a generalized concept of the Hamiltonian will be introduced in Chapter 3.

1.4.2 Elementary properties

Some elementary properties of the configurations in the Kawasaki-Ising model are studied in this subsection, including the dual operator, the shift operator and a counting result.

For a configuration σ on C^d, its *dual configuration* σ^* is defined as the state on C^d such that $\sigma^*(i) = -\sigma(i)$ for $i \in [0, d-1]$. In other words, σ^* is obtained from σ by switching the spin on each vertex of C^d. Note that for $\sigma \in \mathrm{KI}(a, d)$, its dual σ^* belongs to $\mathrm{KI}(b, d)$. Furthermore, we know that $(\sigma^*)^* = \sigma$ holds for any configuration $\sigma \in \mathrm{KI}(a, d)$.

Another important operator acting on σ is the *shift operator* τ, which is defined as follows.

Definition 1.4.1. *Given a configuration σ in $\mathrm{KI}(a, d)$, $\tau(\sigma)$ is a configuration on C^d defined as:*

$$\tau(\sigma)(i) := \sigma_{(i+1)_d} \qquad \forall\ i \in [0, d-1].$$

In other words, $\tau(\sigma) = (\sigma_1, \cdots, \sigma_{d-1}, \sigma_0)$ can be obtained from $\sigma = (\sigma_0, \sigma_1, \cdots, \sigma_{d-1})$ by a shift. One direct observation is that $\tau(\sigma)$ belongs to $\mathrm{KI}(a, d)$ for any σ in $\mathrm{KI}(a, d)$. Furthermore, τ induces an equivalence relationship in $\mathrm{KI}(a, d)$. That is,for any two configurations, $\sigma \sim \sigma'$ if $\sigma = \tau^t(\sigma')$ for some $t \in [1, d]$. Here τ^t means applying the shift operator t times.

The equivalence class of σ, denoted by $[\sigma]$, is called the *shift orbit* of σ. Let $\mathcal{KI}(a, d)$ be the set of all equivalence classes in $\mathrm{KI}(a, d)$. Then $[\sigma] \in \mathcal{KI}(a, d)$. Intuitively, $\mathrm{KI}(a, d)$ consists of the labelled configurations while $\mathcal{KI}(a, d)$ consists of the unlabelled ones.

We end this subsection with a discussion on the following problem: what is the size of $\mathcal{KI}(a, d)$ given two integer parameters (a, d) such that $0 \le a \le d$?

Note that $|\,\mathrm{KI}(a, d)| = \binom{d}{a}$ since there are $\binom{d}{a}$ ways of choosing a vertices from $V(C^d)$ to be the down spins. Let Φ be the group generated by the shift operator τ on $\mathrm{KI}(a, d)$. In other words, Φ is isomorphic to $\mathbb{Z}_d$, which acts naturally on the states in $\mathrm{KI}(a, d)$ as follows:

$$\begin{aligned} \mathbb{Z}_d \times \mathrm{KI}(a, d) &\rightarrow \mathrm{KI}(a, d), \\ (t, \sigma) &\mapsto \tau^t(\sigma). \end{aligned}$$

Note that $|\operatorname{KI}(0,d)| = |\operatorname{KI}(d,d)| = 1$ implies $|\mathcal{KI}(0,d)| = |\mathcal{KI}(d,d)| = 1$.

Let φ be the Euler function. That is, $\varphi(d)$ is the number of integers k in $[1,d]$ such that $\gcd(k,d) = 1$. The following theorem is a well known result in Pólya counting theory; a proof appears in pp.529–530 in [59].

Proposition 1.4.1. *For $1 \le a \le d-1$, we have:*

$$|\mathcal{KI}(a,d)| = \frac{1}{d} \sum_{k|(d,a)} \varphi(k) \binom{d/k}{a/k}.$$

□

When a and d are relatively prime, we have the following simplified form:

$$|\mathcal{KI}(a,d)| = \frac{1}{d}\binom{d}{a}.$$

Clearly, we have $|\mathcal{KI}(1,d)| = |\mathcal{KI}(d-1,d)| = 1$. Note that the above formula shows that the set $\mathcal{KI}(a,d)$ could be very large.

1.4.3 Other visualizations

By interpreting spins in different settings, we can obtain some other visualizations to represent the configurations in $\operatorname{KI}(a,d)$, which is also be denoted by $\operatorname{CONF}(a,b)$.

Let T be a map defined as: $T(+) = 1$ and $T(-) = 0$. With abuse of notation, for each configuration σ in $\operatorname{KI}(a,d)$, we define $T(\sigma)$ as a word ω over alphabet $\{0,1\}$ such that $\omega = T(\sigma_0)T(\sigma_1)\cdots T(\sigma_{d-1})$. Then $T(\sigma)$ belongs to $\mathbb{W}_{a,d}$, the set defined as follows:

$$\mathbb{W}_{a,d} := \{w \in \{0,1\}^d \mid |w|_1 = a\}.$$

In fact, T is a bijection and by which we can virtually identify $\mathbb{W}_{a,d}$ with $\operatorname{KI}(a,d)$. Thus, we obtain another visualization to represent the configurations in $\operatorname{KI}(a,d)$. We should notice that words are a class of objects that have been intensively researched in recent years and have a variety of applications. For more background on words, we recommend Lothaire's book [43].

Another visualization is necklaces, where a configuration in $\mathrm{CONF}(a, b)$ is a cyclic arrangement of a red beads and b black beads. Given a state $\sigma \in \mathrm{KI}(a, a+b)$, by putting a bead on each vertex i in C^{a+b} and coloring it red or black, corresponding to whether $\sigma_i = +1$ or $\sigma_i = -1$, we can associate σ with a necklace. Similarly, from a necklace in $\mathrm{CONF}(a, b)$, we can also construct a state in $\mathrm{KI}(a, d)$. Note that two necklaces are the same if we can rotate one to another. In this setting, an equivalence class of necklaces is a shift orbit in $\mathcal{KI}(a, d)$.

One variant of the above visualization is using polygons, that is, a configuration in $\mathrm{CONF}(a, b)$ is a polygon formed by two different type of sides, say type I and type II. Generally, they are distinguished by length. From a necklace, we can replace each red (resp. black) bead by a side of type I (resp. II) to form a polygon. This visualization plays an important role in Chapter 4.

There is a visualization arising in compute graphics to answer the following problem: how to draw a zig-zag line from $(0, 0)$ to (a, b) on the screen to approximate the "real" line through these two points [11]. We should notice that the screen is represented by the integer lattice $\mathbb{Z}^2$, and one step from (x, y) is either $(x + 1, y)$ (x step) or $(x, y + 1)$ (y step).

Example 1. Some visualizations of the configuration 0101101011 in $\mathrm{CONF}(6, 4)$.

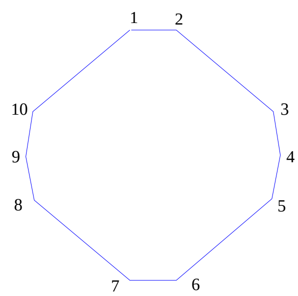

Figure 1.1: A polygon

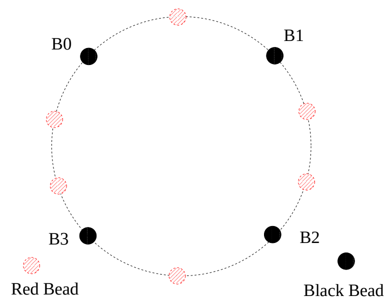

Figure 1.2: A necklace

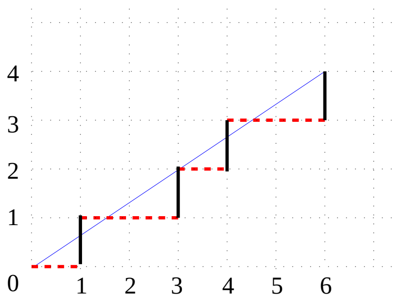

Figure 1.3: A line

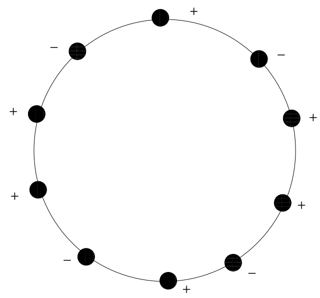

Figure 1.4: Ising Model

We end this subsection with a brief discussion about the connections between musical scales and configurations. For more background and references concerning musical scales, see [17, 16].

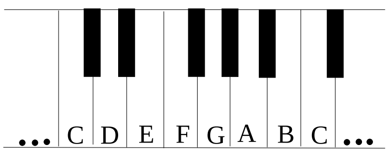

Figure 1.5: Piano keyboard

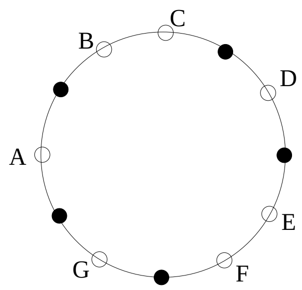

Figure 1.6: CONF(7,5)

Informally speaking, in musical scales, we are interested in how the diatonic set (the white keys on the piano) is embedded in the chromatic scale (all the keys on the piano). Since the arrangement of keys on the piano is periodic, this problem can be reformulated as to arrange 7 white beads (white keys) and 5 black beads (black keys) in a circle. See Figure 1.6 for the configuration in $\mathrm{CONF}(7,5)$ that corresponds to the arrangement of the keys on the piano. By this correspondence, many problems studied in musical scales can be studied in the framework of configurations as well.

Chapter 2

Regular Configurations

In this chapter we study regular configurations. Roughly speaking, the distribution of up spins and down spins in regular configurations is as evenly as possible. Furthermore, we also investigate some properties of these two families of configurations.

2.1 Regularity

In this section, the formal definitions of regular configurations is presented after introducing the degree of regularity.

When $a = 0$, there is only one configuration in $\mathrm{CONF}(a, b)$: the state on C^{a+b} with all vertices associated with down spins or the necklace formed by b black beads. To avoid this trivial case, in the remainder of this chapter, we will assume $a > 0$ and $b > 0$ unless explicitly stated otherwise.

Recall that any configuration σ in $\mathrm{CONF}(a, b)$ is uniquely determined by a pair $(B_0, \mathcal{X})$, where B_0 is the first vertex in the down spin (under the ordering $<$ on $[a+b-1]$) and $\mathcal{X}$ is the characteristic sequence. Then σ in $\mathrm{CONF}(a, b)$ is called *r-regular* for some $r \in [0, b]$ if

$$|x_i + \cdots + x_{i+s-1} - s\frac{a}{b}| < 1 \tag{2.1}$$

for all $i \in [0, b-1]$ and $s \in [1, r]$. Here we use the convention that all configurations are 0-regular and a configuration is at most b-regular.

Note that σ is r-regular for some $r \geq 2$ only if σ is $(r-1)$-regular. Therefore the maximal r such that σ is r-regular but not $(r+1)$-regular will be referred to as the *degree of regularity* and denoted by $\rho(\sigma)$. Note that by definition $0 \leq \rho(\sigma) \leq b$ holds for any configuration $\sigma \in \mathrm{CONF}(a, b)$.

A configuration σ is called *semi-regular* if $\rho(\sigma) > 0$. In particular, we have the following

Definition 2.1.1. *A configuration σ is called* regular *if $\rho(\sigma) = b$. In other words, we have*

$$|x_i + \cdots + x_{i+s-1} - s\frac{a}{b}| < 1 \tag{2.2}$$

for all $i \in [0, b-1]$ and $s \in [1, b]$.

The left side in (2.2) measures the deviation between two quantities: the first one, $x_i + \cdots + x_{i+s-1}$, is the number of up spins between B_i and B_{i+s}; the second one, sa/b, is the expected number of up spins between B_i and B_{i+s} in a random configuration in $\mathrm{CONF}(a, b)$. The smaller value of this deviation (or *discrepancy* as it is sometimes called) would imply the configuration is closer to the "random" one.

By (1.2), the system of inequalities in (2.2) can be simplified as

$$\frac{a}{b}k - 1 < x_i + x_{i+1} + \cdots + x_{i+k-1} < \frac{a}{b}k + 1 \tag{2.3}$$

for all $i \in [0, b-1]$ and $k \in [1, \lfloor b/2 \rfloor]$. Denote this system of inequalities by $\mathrm{Reg}(a, b)$. Then a configuration σ is regular if and only if its characteristic sequence satisfies $\mathrm{Reg}(a, b)$. For simplicity, in the remainder of Part I we shall denote $\lfloor \frac{a}{b} \rfloor$ by $\perp$ and $\lceil \frac{a}{b} \rceil$ by $\top$.

Now we proceed to study some elementary properties of regular configurations.

First note that for any configuration $\sigma \in \mathrm{KI}(a, d)$, $\rho(\sigma) = \rho(\tau(\sigma))$. Furthermore, σ is regular if and only if $\tau(\sigma)$ is regular. Therefore the degree of regularity is invariant under the shift operator, and hence it is also well defined on $\mathcal{KI}(a, a+b)$.

Since $\rho(\sigma)$ is invariant under the shift operator, in the following discussion we will assume $B_0 = 0$ without loss of generality. By this convention, there is a one to one correspondence between configurations in $\mathrm{CONF}(a, b)$ and their characteristic sequences. In the necklace visualization, a configuration σ is represented by the following cyclic sequence:

$$\sigma = (B_0, \underbrace{R, \cdots, R}_{x_0}, B_1, \underbrace{R, \cdots, R}_{x_1}, \cdots, B_{b-1}, \underbrace{R, \cdots, R}_{x_{b-1}}). \tag{2.4}$$

where x_i is the number of red beads between black beads B_i and B_{i+1}. Note that B_j denotes the j-th vertex in the down spin in the Kawasaki-Ising model, but here it denotes the j-th black bead, with abuse of notation.

This gives us another construction of configurations. We can put b black beads in a round and consecutively label them from 0 to $b-1$. Then we put x_i red beads between each pair of black beads B_i and B_{i+1}.

The following lemma provides us with a useful criterion for non-regular configurations.

Lemma 2.1.1. *Given a configuration σ in* $\mathrm{CONF}(a, b)$ *such that $a \geq 2$, if $\rho(\sigma) = r < b$, then there exist i and j such that $B_i < B_j$ and*

$$\mid (x_i + \cdots + x_{i+r}) - (x_j + \cdots + x_{j+r}) \mid \geq 2.$$

Proof. Denoting the sum $x_j + \cdots + x_{j+r}$ by X_j^r, then from the assumption $\rho(\sigma) = r < b$ we can assert that $\mid X_i^r - (r+1)\frac{a}{b} \mid \geq 1$ holds for some i. Now we shall establish the lemma for the case $X_i^r \geq (r+1)\frac{a}{b} + 1$ as a similar argument works for the other case $X_i^r \leq (r+1)\frac{a}{b} - 1$.

Since $X_i^r \geq (r+1)\frac{a}{b} + 1$, it suffices to show that

$$(x_j + \cdots + x_{j+r}) - (r+1)\frac{a}{b} \leq -1$$

holds for some j: Indeed, if the above inequality fails for all j, that is $X_j^r \geq (r+1)a/b$ holds for all $j \in [0, b-1]$, then we have

$$\sum_{j=0}^{b-1} X_j^r \geq b\frac{(r+1)a}{b} + 1 = (r+1)a + 1$$

in view of $X_i^r \geq (r+1)\frac{a}{b} + 1$, a contradiction to

$$\sum_{j=0}^{b-1} X_j^r = \sum_{j=0}^{b-1}(x_j + \cdots + x_{j+r}) = (r+1)(x_0 + \cdots + x_{b-1}) = (r+1)a.$$

□

Now we introduce two useful parameters associated with a configuration.

Definition 2.1.2. *Given a configuration $\sigma \in \mathrm{CONF}(a,b)$ with its characteristic sequence $\mathcal{X} = (x_0, x_1, \cdots, x_{b-1})$. For $1 \leq j \leq b$, we put*

$$\mu_j(\sigma) := \min_{0 \leqslant i \leqslant b-1}\{x_i + x_{i+1} + \cdots + x_{i+j-1}\}; \tag{2.5}$$

$$\xi_j(\sigma) := \max_{0 \leqslant i \leqslant b-1}\{x_i + x_{i+1} + \cdots + x_{i+j-1}\}. \tag{2.6}$$

Here we adopt the convention that $\mu_{-1} = \mu_0 = 0$ and $\xi_{b+1} = a - 1$.

Note that $\mu_j(\sigma)$ is the minimal number of up spins among $j+1$ consecutive down spins in σ, which provides another characterization of regularity.

Lemma 2.1.2. *σ is regular if and only if $1 + \mu_j(\sigma) > (aj)/b$ for $-1 \leqslant j \leqslant b$.*

Proof. "⇒" This direction can be verified directly from the inequalities in $\mathrm{Reg}(a,b)$.

"⇐" In this direction, the left inequalities in $\mathrm{Reg}(a,b)$ are easy. From the assumptions and the fact $x_0 + x_1 + \cdots + x_{b-1} = a$, we have

$$x_i + x_{i+1} + \cdots + x_{i+k-1} \leqslant a - \mu_{b-k} < a - (\frac{a}{b}(b-k) - 1) = 1 + \frac{a}{b}k$$

for all $0 \in [0, b-1]$ and $k \in [1, 1 + \lfloor b/2 \rfloor]$, which completes the proof. $\square$

2.2 Balanced words

In this section, we shall establish the relation between regular configurations and balanced words. Roughly speaking, a word is a configuration written as a cyclic sequence, although it is more common to write 1 for up spin and 0 for down spin. As we see in Section 1.4.3, the set $\mathbb{W}_{a,d}$ formed by the words with length d and containing exactly a 1's is virtually identified with $\mathrm{KI}(a, d)$ by the canonical isomorphism T.

Given a word $w = w_0 w_1 \cdots w_{d-1} \in \{0, 1\}^d$, the *cyclic shift* τ on it is defined as $\tau(w) := w_1 \cdots w_{d-1} w_0$. A *cyclic subword* of w is any length-q prefix of some $\tau^{i-1}(w)$ for i and q in $[1, d]$.

Definition 2.2.1. *A word w is called* balanced *if for any two of its cyclic subwords z and z' with the same length, we have $||z|_i - |z'|_i| \leq 1$ for $i \in \{0, 1\}$.*

Now a configuration is called *balanced* if its representing cyclic sequence is a balanced word. Recall that $[i, i + t]$ denotes the segment $\{i, (i+1)_d, \cdots, (i+t)_d\}$ in $V(C^d)$. For any configuration σ on C^d, a segment of σ, denoted by $\sigma_{[i,i+t]}$, is the word formed by $\sigma_i \sigma_{(i+1)_d} \cdots \sigma_{(i+t)_d}$.

Let $\sigma^{-1}_{[i,i+t]}(s)$ denote $\sigma^{-1}(s) \cap [i, i + t]$ for $s = \pm 1$. In other words, $\sigma^{-1}_{[i,i+t]}(+1)$ contains the vertices in $[i, i + t]$ that are associated with up spins. Note that any cyclic subword can be realized as a segment of the configuration. This implies directly the following

Lemma 2.2.1. *A configuration σ in $\mathrm{KI}(a, d)$ is balanced if and only if*

$$| \, |\sigma^{-1}_{[i,i+t]}(s)| - |\sigma^{-1}_{[j,j+t]}(s)| \, | \leq 1$$

holds for any $s \in \{+1, -1\}$, $i, j \in [0, d-1]$ and $1 \leq t \leq d - 1$.

$\square$

Given a configuration σ in $\mathrm{CONF}(a,b)$, its dual configuration σ^* can be regarded as the state obtained from σ by switching all its spins. In the Kawasaki-Ising model, this means $\sigma^*(i) = -\sigma(i)$ for $i \in [0, d-1]$. Note that σ^* in $\mathrm{CONF}(b,a)$ and $(\sigma^*)^* = \sigma$. Since $|\sigma_{[i,i+t]}^{-1}(-1)| = |\sigma_{[i,i+t]}^{*-1}(+1)|$ holds for any i and t, by the above lemma we have

Corollary 2.2.2. *A configuration σ is balanced if and only if its dual σ^* is balanced.* $\square$

Now we are proceed to establish the connection between regularity and balance.

Theorem 2.2.3. *A configuration σ in $\mathrm{CONF}(a,b)$ is regular if and only if it is balanced.*

Proof. Since the dual operator preserves balance and regularity, in this proof we will assume $a \geq b$ for simplicity.

" $\Leftarrow$ ": This direction is straightforward. For a balanced configuration σ, we assume, for the sake of contradiction, that $\rho(\sigma) = p < b$. By Lemma 2.1.1,

$$(x_j + \cdots + x_{j+p-1}) - (x_i + \cdots + x_{i+p-1}) \geq 2$$

holds for some i and j. Now consider the fragments

$$u = 0_i, \underbrace{1, \cdots, 1}_{x_i}, 0_{i+1}, \cdots, 0_{i+p-1}, \underbrace{1, \cdots, 1}_{x_{i+p-1}}, 0_{i+p}$$

and

$$v = 0_j, \underbrace{1, \cdots, 1}_{x_j}, 0_{j+1}, \cdots, 0_{j+p-1}, \underbrace{1, \cdots, 1}_{x_{j+p-1}}, 0_{j+p}$$

in σ, and construct a new fragment v' by choosing the first $|u|+1$ bits from v and deleting 0_j. Then we have $|u| = |v'|$ and $|u|_0 - |v'|_0 = 2$, a contradiction as required.

" $\Rightarrow$ ": From Lemma 2.2.1, if σ is not balanced, then there exist i, j

in $\mathbb{Z}_d$ and $t \in [1, d-1]$ such that

$$| \, |\sigma^{-1}_{[i,i+t]}(-1)| - |\sigma^{-1}_{[j,j+t]}(-1)| \, | \geq 2.$$

Let u and v denote respectively the segments $\sigma_{[i,i+t]}$ and $\sigma_{[j,j+t]}$. Putting $p := |u|_0$ and $q := |v|_0$, then u can be schematically represented as

$$u = \underbrace{1, \cdots, 1}_{\epsilon_1}, 0_1, \cdots, 0_p, \underbrace{1, \cdots, 1}_{\epsilon_2}$$

with $0 \leq \epsilon_1 \leq x_0$ and $0 \leq \epsilon_2 \leq x_p$. Similarly, we have

$$v = \underbrace{1, \cdots, 1}_{\epsilon'_1}, 0_l, \cdots, 0_{l+q-1}, \underbrace{1, \cdots, 1}_{\epsilon'_2}$$

with $0 \leq \epsilon'_1 \leq x_{l-1}$ and $0 \leq \epsilon'_2 \leq x_{l+q-1}$. Since $|u| = |v| = t+1$, clearly we have

$$\epsilon_1 + x_1 + \cdots + x_{p-1} + \epsilon_2 + p = \epsilon'_1 + x_l + \cdots + x_{l+q-2} + \epsilon'_2 + q = t + 1. \quad (2.7)$$

Assume without loss of generality that $p - q \geq 2$ holds. Then from equation (2.7) and the constraints of ϵ and ϵ', this implies

$$(x_{l-1} + x_l + \cdots + x_{l+q-1}) - (x_1 + \cdots + x_{p-1}) \geq 2. \quad (2.8)$$

On the other hand, since σ is regular, we have

$$\begin{aligned} (x_{l-1} + \cdots + x_{l+p-1}) - (x_1 + \cdots + x_{q-1}) &< (q+1)\frac{a}{b} + 1 - [(p-1)\frac{a}{b} - 1] \\ &= (q + 2 - p)\frac{a}{b} + 2 \\ &\leq 2, \end{aligned}$$

a contradiction as required. Note that in the last step of the above inequalities we also use the assumption that $q + 2 \leq p$. $\square$

By Theorem 2.2.3, we have the following two further properties of regular configurations: the first one derives from Corollary 2.2.2 and the second one follows from the fact that the balanced word with a given number of 1s and 0s is unique (up to shifting) [37].

Theorem 2.2.4. *A configuration σ in* $\mathrm{CONF}(a, b)$ *is regular if and only if its dual configuration σ^* is regular.* □

Theorem 2.2.5. *The regular configurations in* $\mathrm{CONF}(a, b)$ *are unique up to shifting.* □

Let us remark that there exist many well well-known algorithms to construct regular configurations [11, 23].

2.3 Properties

In the last section of this chapter, we collect some properties of regular configurations.

2.3.1 Self-similarity

Recall that for a given configuration $\sigma \in \mathrm{CONF}(a, b)$, its characteristic sequence $\mathcal{X} = (x_0, \cdots, x_{b-1})$ is given by $x_i = (B_{(i+1)_b} - B_i - 1)_{(a+b)}$ for $i \in [0, b-1]$. Another convention we have adopted is $\perp = \lfloor \frac{a}{b} \rfloor$ and $\top = \lceil \frac{a}{b} \rceil$.

Similarly to the definition of balance over $\{0, 1\}$, a word w over $\{\top, \perp\}$ is called *balanced* if and only if $||z|_\top - |z'|_\top| \leq 1$ for any two cyclic subwords z and z'. Here $|z|_\top$ and $|z|_\perp$ denote respectively the number of the occurrences of $\top$ and $\perp$ in z. Note that if $\top = \perp$, then all words over $\{\top, \perp\}$ are balanced.

The following theorem gives another characterization of regular configurations.

Theorem 2.3.1. *A configuration $\sigma \in$ $\mathrm{CONF}(a, b)$ is regular if and only if its characteristic sequence is a balanced word over $\{\top, \perp\}$.*

Proof. Clearly, if σ is regular, then its characteristic sequence is a word over $\{\top, \bot\}$, which will be denoted by w in this proof, i.e., $w = x_0x_1 \cdots x_{b-1}$.

For any two cyclic subwords z, z' of length s in w, we know

$$z = x_ix_{i+1} \cdots x_{i+s-1} \quad \text{and} \quad z' = x_jx_{j+1} \cdots x_{j+s-1}$$

for some $i, j \in [0, b-1]$, where the subscripts are calculated modulo b. Since

$$||z|_\top - |z'|_\top| = |(x_i + x_{i+1} + \cdots + x_{i+s-1}) - (x_j + x_{j+1} + \cdots + x_{j+s-1})|,$$

we can assert that $||z|_\top - |z'|_\top|| \leq 1$ holds for each pair of subwords of length s if and only if we have $\xi_s - \mu_s \leq 1$. As this assertion holds for all $s \in [1, b]$, the proof is completed. $\square$

For regular configurations, their characteristic configurations are well defined. Together with Theorem 2.2.3, the above theorem has the following corollary.

Corollary 2.3.2. *A configuration in* $\mathrm{CONF}(a, b)$ *is regular if and only if its characteristic configuration is regular.* $\square$

2.3.2 Symmetry

In this section, we are going to study the symmetry of the regular configuration in $\mathrm{CONF}(a, b)$.

Recall that the cyclic group generated by the shift operator τ on $\mathrm{CONF}(a, b)$ is denoted by Φ. For any $\sigma \in \mathrm{CONF}(a, b)$, let Φ_σ denote the stationary subgroup of Φ whose elements fix σ, and let $\mathrm{Orb}(\sigma)$ be the orbit of σ under the action of Φ. In other words, we have

$$\Phi_\sigma = \{\tau^t \mid \tau^t(\sigma) = \sigma \text{ and } 0 \leq t \leq a + b - 1\},$$

and

$$\mathrm{Orb}(\sigma) = \{\tau^t(\sigma) \mid 0 \leq t \leq a + b - 1\}.$$

Let κ be the minimal positive integer such that $\tau^{\kappa}(\sigma) = \sigma$. Then Φ_σ is a cyclic subgroup of $\mathbb{Z}_{a+b}$ generated by κ. Furthermore, we have $|\Phi_\sigma| = (a+b)/\kappa$ and $|\operatorname{Orb}(\sigma)| = \kappa$.

The *symmetry degree* of a configuration σ, $\chi(\sigma)$, is defined as $|\Phi_\sigma|$. By this definition, $\chi(\sigma) \in [1, a+b]$ holds for any $\sigma \in \mathrm{CONF}(a,b)$. Indeed, by Lagrange's Theorem, it is not difficult to see that $\chi(\sigma) \leq \gcd(a,b)$ holds for any configuration $\sigma \in \mathrm{CONF}(a,b)$. Hence we have the following

Definition 2.3.1. *A configuration σ in* $\mathrm{CONF}(a,b)$ *is called* symmetric *if $\chi(\sigma) = \gcd(a,b)$.*

In other words, a configuration is called symmetric if it has the maximal possible symmetry degree.

Example 2. Figure 1.1 and Figure 2.2 show two symmetric configurations for $\mathrm{CONF}(6,4)$. On the other hand, Figure 2.1 presents an example of nonsymmetric configurations. Note that the configuration in Figure 2.2 is symmetric but not regular.

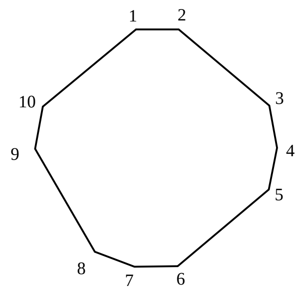

Figure 2.1: Nonsymmetric

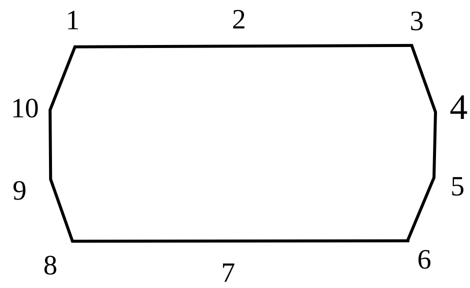

Figure 2.2: Symmetric

As Fig 1.1 also presents a symmetric configuration in $\mathrm{CONF}(6,4)$, the above example shows that the symmetric configurations in $\mathrm{CONF}(a,b)$ are generally not unique, even under the action of τ. Therefore, symmetry does not imply regularity. But the converse is true, as the following theorem implies.

Theorem 2.3.3. *Regular configurations are symmetric.*

Proof. We shall prove $\chi(\sigma) = \gcd(a, b)$ for the regular configuration $\sigma \in \mathrm{CONF}(a, b)$ by a straightforward induction on b.

Step 1: The base case is $a = bt$, which includes $b = 1$, is straightforward as the regular configuration σ in $\mathrm{CONF}(a, b)$ is characterized by $(t, \cdots, t)$.

Step 2: Now assume the theorem fails for some $\mathrm{CONF}(a, b)$ and let b be minimal such that the regular configuration σ in $\mathrm{CONF}(a, b)$ satisfies $|\Phi_\sigma| < \gcd(a, b)$. From step 1 we have $a = bt + r$ for some $t \in \mathbb{N}$ and integer $0 < r < b$. Denoting the characteristic sequence of σ by $\mathcal{X} = \{x_0, \cdots, x_{b-1}\}$, then the configuration σ' with the characteristic sequence $\mathcal{X} - t$ is a regular configuration in $\mathrm{CONF}(r, b)$ and $\chi(\sigma) = \chi(\sigma')$. Since $\gcd(a, b) = \gcd(r, b)$, σ' is not symmetric, a contradiction to the minimality of b. $\square$

One direct consequence of the above theorem is the following corollary, which counts the number of regular configurations in $\mathrm{CONF}(a, b)$.

Corollary 2.3.4. *For a regular configuration σ in $\mathrm{CONF}(a, b)$, there are exactly $(a + b)/\gcd(a, b)$ configurations in $\mathrm{Orb}(\sigma)$.* $\square$

Since $\gcd(a, b) = \gcd(a, a + b)$, the above result shows that there are exactly $d/\gcd(a, d)$ regular configurations in $\mathrm{KI}(a, d)$, and they form a unique shift orbit in $\mathcal{KI}(a, d)$.

2.3.3 Maximally even sets

In this section, we shall show the equivalence between regular configurations and maximally even sets (ME), a concept in musical scales that has been intensively studied [17].

Given a configuration $\sigma \in \mathrm{CONF}(a, b)$ and its characteristic sequence $(x_0, x_1, \cdots, x_{b-1})$, the spectrum of σ is defined as $\Xi := \{\Xi_1, \Xi_2, \cdots, \Xi_{b-1}\}$, where Ξ_r $(1 \le r \le b)$ is a multiset defined as

$$\Xi_r := \{x_i + x_{i+1} + \cdots + x_{i+r-1} \mid 0 \le i \le b - 1\}.$$

Let $\underline{\Xi_r}$ be the underlying set of Ξ_r. Then we have the following definition of ME, which is slightly different from the original one [17], but they are essentially the same.

Definition 2.3.2. *Let Ξ be the spectrum of σ. Then σ is called* maximally even *if and only if each $\underline{\Xi_r}$ $(1 \le r \le b)$ contains either one integer or two consecutive integers.*

Note that, for $1 \le r \le b$, μ_r and ξ_r, which are given as Definition 2.1.2, belong to Ξ_r. From the above definition, if σ is maximally even, $\xi_r - \mu_r \le 1$ for $1 \le r \le b$. This leads to the following theorem.

Theorem 2.3.5. *A configuration σ in* CONF(a, b) *is regular if and only if it is maximally even.*

Proof. "$\Rightarrow$" If σ is regular, then its characteristic sequence satisfies Reg(a,b). This implies

$$\mu_r, \xi_r \in \left(\frac{ar}{b} - 1, \frac{ar}{b} + 1\right) \quad \text{for } 1 \le r \le b.$$

Therefore $\xi_r - \mu_r \le 1$, since both of them are integers. This means that $\underline{\Xi_r}$ contains either an integer or two consecutive integers.

"$\Leftarrow$" We will prove this direction by contradiction. If σ is not regular, then from Lemma 2.1.1, there exists $r \in [1, b]$ such that $\xi_r - \mu_r \ge 2$, a contradiction to the fact that σ is maximally even. $\square$

Chapter 3

Ground States

In this chapter, we show a connection between the energy of a configuration and its degree of regularity. Loosely speaking, the higher its degree of regularity, the lower its energy. This fact leads to a characterization of regular configurations by ground states, the states with the minimum energy. Furthermore, it also presents a new interpretation of balanced words.

In the second part of this chapter, we study the dynamics of the Kawasaki-Ising model. Roughly speaking, for any non-regular configurations, a particular path in the state space that connects it to a regular configuration is investigated. To this end, we develop some stability lemmas for the structure of non-regular configurations. One byproduct here is another proof of the equivalence between regular configurations and ground states.

3.1 Hamiltonian

In this section, we introduce the Hamiltonian on the Kawasaki-Ising model, which can be regarded as a generalization of of that on the Ising model.

For the Ising model on C^d, the Hamiltonian for a given configuration

$\sigma = (\sigma_0, \cdots, \sigma_{d-1})$ is defined in Section 1.4 as

$$H_1(\sigma) = J_1 \sum_{i=0}^{d-1} \sigma_i \sigma_{i+1}. \tag{3.1}$$

Here J_1 is a positive constant, i.e., the interactions between the neighboring spins are antiferromagnetic.

But for the Kawasaki-Ising model, we consider long-range interactions as well. That is, the actions between σ_i and σ_{i+j} for all j. To measure such interactions, we introduce the generalized Hamiltonian H_p as follows.

Definition 3.1.1. *For $1 \leq p \leq d$, the p-th Hamiltonian component H_p on a configuration σ in* KI(a, d) *is defined as*

$$H_p(\sigma) := J_p \sum_{i=0}^{d-1} \sigma_i \sigma_{i+p}. \tag{3.2}$$

Note that when $p = 1$, $H_p(\sigma)$ is exactly the Hamiltonian of σ when it is regarded as the Ising model. In this chapter, we will let

$$J_p = \left(\frac{1}{2d}\right)^p$$

for $p \in [1, d]$. Actually , as shown in [22, 23], it suffices to let J_p satisfy certain "convexity" conditions to define the Hamiltonian in this model. For this thesis, we fix these parameters as above for simplicity.

Definition 3.1.2. *The* Hamiltonian *of a configuration σ, denoted by $H(\sigma)$, is defined as the sum of its p-th Hamiltonian components:*

$$H(\sigma) := \sum_{p=1}^{d} H_p(\sigma). \tag{3.3}$$

A *ground state* in KI(a, d) is a configuration that has the minimum Hamiltonian over all configurations in KI(a, d). Note that Hamiltonian induces an pre-ordering on the configurations in $\mathrm{KI}(a,d)$. More precisely, if $H(\sigma) \geq H(\phi)$ for σ and ϕ in $\mathrm{KI}(a,d)$, then we say $\sigma \succeq_E \phi$. It is clear that $\succeq_E$ satisfies transitivity and $\sigma \succeq_E \sigma$ for any $\sigma \in \mathrm{KI}(a,d)$. But antisymmetry does not always hold, as the following example shows.

Example 3. Consider two configurations $\sigma, \phi \in \mathrm{KI}(4,12)$, where $\sigma^{-1}(+1) = \{0,1,4,6\}$ and $\phi^{-1}(+1) = \{0,1,3,7\}$. Clearly $H(\sigma) = H(\phi)$. But σ and ϕ are not equivalent, even under the action of shifting and reflecting.

We can give another interpretation of the pre-ordering $\succeq_E$. For a configuration σ, consider the *Hamiltonian vector* associated with it as follows.

$$\vec{H}(\sigma) = \langle H_1(\sigma), \cdots, H_d(\sigma) \rangle.$$

Then $\succeq_E$ is induced by the lexicographic ordering of $\mathbb{R}^d$. More precisely, we have the following lemma.

Lemma 3.1.1. *For any two configurations σ and ϕ in $\mathrm{KI}(a,d)$, $\sigma \succ_E \phi$ if and only if there exists an integer $l \in [1,d]$ such that*

$$H_i(\sigma) = H_i(\phi) \text{ for } 1 \leq i \leq l-1 \quad \text{and} \quad H_l(\sigma) > H_l(\phi). \tag{3.4}$$

Proof. From

$$|\sum_{i=0}^{d-1} \sigma_i \sigma_{i+p}| \leq d \quad \text{and} \quad J_p = \left(\frac{1}{2d}\right)^p,$$

we have

$$|H_p| \leq \frac{1}{2}\left(\frac{1}{2d}\right)^{p-1} \tag{3.5}$$

holds for any $p \in [1,d]$. Therefore,

$$|\sum_{p=l+1}^{d} H_p| < J_l = \left(\frac{1}{2d}\right)^l \tag{3.6}$$

holds for any $l \in [1,d]$. This implies the lemma. $\square$

Now we associate each configuration σ in $\mathrm{KI}(a,d)$ with a new family of parameters, which can used to replace the role of H_p in characterizing $\succ_E$ but is better for calculations.

For each integer $p \in [1, d]$, let $N_p(\sigma)$ consists of the ordered pair $(i, i+p)$ such that $\sigma_i = \sigma_{i+p} = -1$. Formally, we have

$$N_p(\sigma) = \{(i, i+p) \mid 0 \leq i < d,\ \sigma_i = \sigma_{i+p} = -1\}.$$

Furthermore, for each integer $p \in [1, d]$, let $I_p(\sigma)$ be defined as

$$I_p(\sigma) := \frac{1}{4} \sum_{i=0}^{d-1} (1 - \sigma_i)(1 - \sigma_{i+p}). \tag{3.7}$$

Note that $(1 - \sigma_i)(1 - \sigma_{i+p})$ is equal to 4 when $\sigma_i = \sigma_{i+p} = -1$ and is equal to 0 in the other cases. Therefore $I_p(\sigma)$ counts the number of pairs (σ_i, σ_{i+p}) such that $\sigma_i = \sigma_{i+p} = -1$. In other words, we have the following

Lemma 3.1.2. *For any integer $p \in [1, d]$, we have $I_p(\sigma) = |N_p(\sigma)|$.* $\square$

Given a configuration $\sigma \in \mathrm{CONF}(a, b)$, $p \in [1, a+b]$ and $t \in [1, b]$, let $\vartheta_p^t(\sigma)$ denote the multiplicity of p in $t + \Xi_t(\sigma)$. Then the following lemma presents another formula to calculate $I_p(\sigma)$ by its spectrum.

Lemma 3.1.3. *Given a configuration $\sigma \in \mathrm{CONF}(a, b)$ and $p \in [1, a+b]$, we have*

$$I_p(\sigma) = \sum_{i=1}^{b} \vartheta_p^i(\sigma).$$

Proof. The proof is straightforward. Let $d = a + b$. If $\sigma_j = \sigma_{(j+p)_d} = -1$ for some $j \in [0, d-1]$, then $\sigma_j = B_t$ while $\sigma_{(j+p)_d} = B_{(t+s)_b}$ for some $t \in [0, b-1]$ and $s \in [1, b-2]$. In other words,

$$p = s + (x_t + x_{t+1} + \cdots + x_{t+s-1}).$$

Therefor we have

$$p \in s + \Xi_s(\sigma).$$

On the other hand, for each occurrence of p in $s + \Xi_s(\sigma)$, we can find a different pair $(j, j+p)$ such that $\sigma_j = \sigma_{(j+p)_d} = -1$. This completes the proof. $\square$

Lemma 3.1.4. *For any integer $p \in [1, d]$ and $\sigma \in \mathrm{KI}(a, d)$, we have*

$$\sum_{i=0}^{d-1} \sigma_i \sigma_{i+p} = a + 4I_p(\sigma) - 3b, \tag{3.8}$$

where $b = d - a$.

Proof. It is clear that

$$\sum_{i=0}^{d-1} \sigma_i = a - b.$$

Together with equation (3.7), we have

$$\begin{aligned}
4I_p(\sigma) &= \sum_{i=0}^{d-1}(1-\sigma_i)(1-\sigma_{i+p}) \\
&= \sum_{i=0}^{d-1}(1-\sigma_i-\sigma_{i+p}) + \sum_{i=0}^{d-1}\sigma_i\sigma_{i+p} \\
&= (a+b) - 2\sum_{i=0}^{d-1}\sigma_i + \sum_{i=0}^{d-1}\sigma_i\sigma_{i+p} \\
&= (3b-a) + \sum_{i=0}^{d-1}\sigma_i\sigma_{i+p}.
\end{aligned}$$

Rearrange the items in both sides give us (3.8). $\square$

From the above lemma and the fact that J_p is a positive constant, we can assert that for any two configurations $\sigma, \phi \in \mathrm{KI}(a, d)$ and any $p \in [1, d]$, $H_p(\sigma) > H_p(\phi)$ holds if and only if $I_p(\sigma) > I_p(\phi)$ holds. Together with Lemma 3.1.1, this assertion implies the following useful criterion.

Corollary 3.1.5. *For any two configurations σ and ϕ in $\mathrm{KI}(a, d)$, $\sigma \succ_E \phi$*

if and only if there exists an integer $l \in [1, d]$ such that

$$I_i(\sigma) = I_i(\phi) \text{ for } 1 \leq i \leq l-1 \quad \text{and} \quad I_l(\sigma) > I_l(\phi).$$

□

In other words, let

$$I(\sigma) := \langle I_1(\sigma), \cdots, I_d(\sigma) \rangle$$

be a vector in $\mathbb{N}^d$. Then $\sigma \succ_E \phi$ if and only if $I(\phi) <_L I(\sigma)$. Here $<_L$ denotes the lexicographic ordering.

In the following two lemmas, we will show that $H(\sigma)$ is invariant under the dual operator $*$ and the shift operator τ.

Lemma 3.1.6. *For any configuration σ in $\mathrm{KI}(a, d)$ and its dual σ^*, we have $H(\sigma) = H(\sigma^*)$.*

Proof. Recall that σ^* is a configuration in $\mathrm{KI}(d-a, d)$ defined as $(\sigma^*_0, \cdots, \sigma^*_{d-1})$, where $\sigma^*_i = -\sigma_i$. Thus

$$\sigma^*_i \sigma^*_{i+p} = \sigma_i \sigma_{i+p}$$

holds for any integer $p \in [1, d]$. This means that

$$H_p(\sigma) = H_p(\sigma^*)$$

holds for any p, a sufficient condition for $H(\sigma) = H(\sigma^*)$. □

Lemma 3.1.7. *If $\sigma \sim \phi$, then $H(\sigma) = H(\phi)$.*

Proof. From the definition of the shift operator τ, we have

$$\sum_{i=0}^{d-1} \sigma_i \sigma_{i+p} = \sum_{i=0}^{d-1} \sigma_{i+1} \sigma_{i+p+1} = \sum_{i=0}^{d-1} \tau(\sigma)_i \tau(\sigma)_{i+p} \tag{3.9}$$

for any $p \in [1, d]$. This implies that

$$H_p(\sigma) = H_p(\tau(\sigma))$$

holds for any $p \in [1, d]$. Therefore $H(\sigma) = H(\tau(\sigma))$. Since $\sigma \sim \phi$ if and only if $\phi = \tau^j(\sigma)$ for some j, we conclude that $H(\phi) = H(\sigma)$. $\square$

3.2 Ground states

In this section, we investigate the connections between regularity and energy. Recalling that $\rho(\sigma)$, the degree of the configuration σ, is defined to be r if σ is r-regular but not $(r+1)$-regular. As the following theorem shows, it is closely linked with the energy of σ.

Theorem 3.2.1. *Suppose that σ, ϕ are two configurations in* $\mathrm{CONF}(a, b)$ *such that $\rho(\sigma) > \rho(\phi)$. Then $H(\sigma) < H(\phi)$.*

Proof. It suffices to show that if σ is r-regular while ϕ is $(r-1)$-regular but not r regular for some $r \in \mathbb{N}^+$, then $H(\sigma) < H(\phi)$, which can be established by the following straightforward verification.

Put
$$\alpha := \left\lfloor \frac{ra}{b} \right\rfloor \quad \text{and} \quad \beta := \left\lceil \frac{ra}{b} \right\rceil.$$
Without loss of generality, we will assume that $\alpha \neq \beta$ in this proof. Since σ is r-regular, we have $\Xi_r^\sigma = \{\alpha^p, \beta^q\}$, where p and q are respectively the multiplicity of α and β in Ξ_r^σ. On the other hand, let p' (resp. q') denote the multiplicity of α (resp. β) in Ξ_r^ϕ. Then $\Xi_r^\phi - \{\alpha^{p'}, \beta^{q'}\}$ contains t elements (counting multiplicity) for a positive integer t, since ϕ is not r-regular. Now we show that $H(\sigma) < H(\phi)$ by considering the following two cases.

Case I: The minimal element in Ξ_r^ϕ, say λ, is smaller than α. Now we can assert that $I_{\lambda+r}(\phi) > I_{\lambda+r}(\sigma)$, while $I_s(\sigma) = I_s(\phi)$ for $1 \leq s < \lambda + r$, from Lemma 3.1.3 and the following facts:

(1): $\vartheta_j^i(\sigma) = \vartheta_j^i(\phi)$ if $i < r$; (Note that if $i \leq r-1$, then $\Xi_i(\sigma) = \Xi_i(\phi)$ since σ and ϕ are both $(r-1)$-regular.)

(2): $\vartheta_{\lambda+r}^r(\sigma) = 0$ and $\vartheta_{\lambda+r}^r(\phi) > 0$; (Clearly, $\lambda \notin \Xi_r(\sigma)$ and $\lambda \in \Xi_r(\phi)$.)

(3): $\vartheta_j^r(\sigma) = \vartheta_j^r(\phi) = 0$ for $j < \lambda + r$; (If $j < \lambda + r$, then $j - r \notin \Xi_r^\phi \cup \Xi_r^\sigma$.)

(4): $\vartheta_j^i(\sigma) = \vartheta_j^i(\phi) = 0$ if $i > r$ and $j \leq \lambda + r$. (If $i > r$ and $j \leq \lambda + r$, then $j \notin i + \Xi_i(\phi)$ and $j \notin i + \Xi_i(\sigma)$, since $j < \lambda + i$ and $\lambda \leq k$ for any $k \in \Xi_i^\phi \cup \Xi_i^\sigma$.)

Case II: The t elements in $\Xi_r^\phi - \{\alpha^{p'}, \beta^{q'}\}$ are greater than β. In other words, the arithmetic average of such t elements, denoted by c, is greater than β. Now we have

$$p\alpha + q\beta = p'\alpha + q'\beta + tc, \tag{3.10}$$

and

$$p + q = p' + q' + t, \tag{3.11}$$

since the sum of the elements in $\Xi_r(\sigma)$ is equal to that of $\Xi_r(\sigma)$ and they have the same cardinality.

Combining (3.11) and (3.10), we obtain

$$(p - p')(\alpha - \beta) = t(c - \beta).$$

Together with the assumption that $\alpha < \beta < c$ and $t \geq 1$, the above equality implies that $p < p'$. Now we conclude that $I_{\alpha+r}(\phi) > I_{\alpha+r}(\sigma)$, while $I_s(\sigma) = I_s(\phi)$ for $1 \leq s < \alpha + r$, from Lemma 3.1.3 and the following facts:

(1): $\vartheta_j^i(\sigma) = \vartheta_j^i(\phi)$ if $i < r$; (This is the same as that in Case I.)

(2): $\vartheta_{\alpha+r}^r(\sigma) < \vartheta_{\alpha+r}^r(\phi)$; (Note that $\vartheta_{\alpha+r}^r(\sigma) = p$ and $\vartheta_{\alpha+r}^r(\phi) = p'$.)

(3): $\vartheta_j^r(\sigma) = \vartheta_j^r(\phi) = 0$ for $j < \alpha + r$; (If $j < \alpha + r$, then $j - r \notin \Xi_r^\phi \cup \Xi_r^\sigma$.)

(4): $\vartheta_j^i(\sigma) = \vartheta_j^i(\phi) = 0$ if $i > r$ and $j \leq \alpha + r$. (This can be proved by a similar argument to that in Case I.)

Since in both cases, there exists $l \in [1, a+b]$ such that $I_i(\sigma) = I_i(\phi)$ for $i \in [1, l]$ while $I_{l+1}(\sigma) < I_{l+1}(\phi)$, the theorem follows from Corollary 3.1.5 $\square$

Note that the above theorem provides us with a new characterization of regular configurations.

Theorem 3.2.2. *For any configuration σ in* $\mathrm{CONF}(a, b)$*, σ is regular if and only if it is a ground state.*

Proof. This follows from Theorem 3.2.1 and the fact that the regular configuration in $\mathrm{CONF}(a, b)$ is the only one in $\mathrm{CONF}(a, b)$ that has the maximal degree of regularity. $\square$

Together with Theorem 2.2.3, the above theorem implies the following corollary.

Corollary 3.2.3. *A configuration σ in* $\mathrm{CONF}(a, b)$ *is balanced if and only if it is a ground state.* $\square$

Thus, we obtain a new characterization of balanced words, a well studied object in words, by a concept in statistical mechanics.

3.3 Dynamics

In this section, we discuss some dynamic aspects of the Kawasaki-Ising model on cycles. To this end, we introduce the concept of state graph.

3.3.1 State graph

We begin with recalling some notation defined in Chapter 1. Given $i, j \in V(C^d) = \{0, 1, \cdots, d-1\}$, the interval $[i, j]$ on the cycle C^d is defined to be the path $i, (i+1)_d, \cdots, j$, and the number of edges in this path is written as $|j-i|_L$.

One important operator defined in the Kawasaki-Ising model is the switch operator $\mathcal{S}$, which acts on states in $\mathrm{CONF}(a, b)$ by switching the spins on some interval $[i, j]$ of C^{a+b}. When $j = i+1$, it simply switches one pair of neighboring spins. More precisely, we have the following definition.

Definition 3.3.1. *Suppose that σ is a configuration in* $\mathrm{CONF}(a,b)$ *and* $i, j \in V(C^{a+b})$. *Then $\mathcal{S}_{i,j}(\sigma)$ is a configuration in* $\mathrm{CONF}(a,b)$ *defined as*

$$\mathcal{S}_{i,j}(\sigma)(t) := \begin{cases} \sigma_{(j-|t-i|_L)_{a+b}} & \text{if } t \in [i,j], \\ \sigma_t & \text{if } t \notin [i,j]. \end{cases}$$

Denoting $\mathcal{S}_{i,j}(\sigma)$ by ϕ, then it is clear that $\phi_i = \sigma_j$ and $\phi_j = \sigma_i$. When the subindex is not important or is clear from the context, we will abbreviate $\mathcal{S}_{i,j}(\sigma)$ to $\mathcal{S}(\sigma)$. The above definition is further illustrated by the following example.

Example 4. Considering the configuration

$$\sigma = (+,-,+,+,-,-,+,+,+)$$

on the cycle $C = \{0, 1, \cdots, 8\}$. Then we have

$$\mathcal{S}_{0,4}(\sigma) = (-,+,+,-,+,-,+,+,+).$$

Denoting a set of x up spins by $\boxed{x}$, in this thesis we will mainly use the the operator $\mathcal{S}$ in the following scenario. Suppose the configuration σ contains the following segment of spins:

$$-\boxed{x_j}-\boxed{x_{j+1}}-\cdots-\boxed{x_{j+t-1}}-\boxed{x_{j+t}}-,$$

where the down spin in the right of $\boxed{x_j}$ is in position u and the first up spin in $\boxed{x_{j+t}}$ is in position v. Then $\phi = \mathcal{S}_{u,v}(\sigma)$ has the same spin structure as that of σ except for replacing the above fragment by the following one:

$$-\boxed{x_j+1}-\boxed{x_{j+t-1}}-\cdots-\boxed{x_{j+1}}-\boxed{x_{j+t}-1}-.$$

In other words, if σ is characterized by the sequence

$$(x_0, \cdots, x_j, x_{j+1}, \cdots, x_{j+t-1}, x_{j+t}, \cdots, x_{b-1}),$$

then the characterizing sequence of $\phi = \mathcal{S}_{u,v}(\sigma)$ is

$$(x_0, \cdots, x_j + 1, x_{j+t-1}, x_{j+t-2}, \cdots, x_{j+2}, x_{j+1}, x_{j+t} - 1, \cdots, x_{b-1}).$$

Similarly, the characterizing sequence of $\phi' = \mathcal{S}_{u-1,v-1}(\sigma)$ is

$$(x_0, \cdots, x_j - 1, x_{j+t-1}, x_{j+t-2}, \cdots, x_{j+2}, x_{j+1}, x_{j+t} + 1, \cdots, x_{b-1}).$$

Note that both ϕ and ϕ' can be obtained from σ by applying the switch operator once.

It is easy to see that $\mathcal{S}$ is an involution. More precisely, if we apply $\mathcal{S}$ on σ twice, then we get σ again. Now we use this operator to define a graph on the set of all states in $\mathrm{CONF}(a, b)$.

Definition 3.3.2. *The* state graph $\mathcal{G}_{a,b}$ *of the Kawasaki-Ising model* $\mathrm{CONF}(a, b)$ *is given as follows: the vertex set consists of the states in* $\mathrm{CONF}(a, b)$*, and two vertices* (σ, ϕ) *are adjacent if and only if* $\mathcal{S}(\sigma) = \phi$*.*

It is straightforward to verify that $\mathcal{G}_{a,b}$ is connected and its maximal degree is bounded by $\binom{a+b}{2}$. In the remainder of this section, we will investigate some other aspects of state graphs.

3.3.2 Monotone paths

The following theorem is the main result of this section, which claims that non-regular configurations can "evolve" to regular configurations via an "energy decreasing " path in the state graph.

Theorem 3.3.1. *Suppose that σ is a non-regular configuration in* $\mathrm{CONF}(a, b)$*. Then there exists a path $P = v_0 \cdots v_t$ in $\mathcal{G}_{a,b}$ such that $v_0 = \sigma$, $v_t = \phi$ for a regular configuration in* $\mathrm{CONF}(a, b)$ *and $H(v_k) > H(v_{k+1})$ for each $k \in [0, t-1]$,*

Proof. From Theorem 2.2.4 and Lemma 3.1.6, the duality operator $*$ preserves regularity and the Hamiltonian. Therefore we can assume $a \geq b$

in this proof since otherwise we can replace ϕ by its dual configuration ϕ^*.

It suffices to show that if σ is non-regular, then there exists ϕ such that $(\sigma, \phi) \in E(\mathcal{G}_{a,b})$ and $H(\sigma) > H(\phi)$. We have divided the proof of this observation into a sequence of lemmas (Lemma 3.3.3, 3.3.4 and 3.3.6) by considering three possible values of $\rho(\sigma)$: $\rho(\sigma) = 0$, $\rho(\sigma) = 1$ and $\rho(\sigma) = h - 1$ for $3 \leq h \leq b$. Their proofs are quite involved and will be presented in Subsection 3.3.3. $\square$

Informally speaking, associating each vertex $u \in \mathcal{G}_{a,b}$ with its energy, the Hamiltonian $H(u)$, Theorem 3.2.2 shows that the regular configurations in $\mathrm{CONF}(a, b)$ are the vertices that have the "globally" minimal energy. Furthermore, Theorem 3.3.1 implies that there are no other vertices with the "locally" minimal energy, i.e., given any non-regular configuration σ in $\mathcal{G}_{a,b}$, there exits a configuration in its neighbors that has higher energy.

The following corollary is a direct consequence of the above theorem.

Corollary 3.3.2. *For $\sigma \in \mathrm{CONF}(a, b)$, if $H(\sigma)$ is minimal, then σ is regular.*

$\square$

The above corollary also leads to an alternative proof of Theorem 3.2.2.

Another proof of Theorem 3.2.2. From Corollary 3.3.2, it is sufficient to show that $H(\sigma)$ is minimal if σ is regular. If this fails, then there exists a configuration ϕ such that $H(\phi)$ is minimal and $H(\phi) < H(\sigma)$. On the other hand, we can assert that ϕ is regular from Corollary 3.3.2. By Theorem 2.2.5 and Lemma 3.1.7, this implies that $H(\phi) = H(\sigma)$, a contradiction. $\square$

3.3.3 Technical lemmas

In this subsection we will prove the lemmas used in the proof of Theorem 3.3.1. Throughout this subsection, we will use the convention that $a > b$. We begin with the case that $\rho(\sigma) = 0$.

Lemma 3.3.3. *Suppose that σ is a non-regular configuration in* $\mathrm{CONF}(a, b)$ *with $\rho(\sigma) = 0$. Then there exists a configuration ϕ such that $(\sigma, \phi) \in E(\mathcal{G}_{a,b})$ and $H(\sigma) > H(\phi)$.*

Proof. Recall that $d_c(x_i, x_j)$ for x_i, x_j in the characteristic sequence of σ is defined as $\min\{(i-j)_b, (j-i)_b\}$. Among all pairs (i, j) satisfying the inequality $|x_i - x_j| \geq 2$, consider one pair such that $t = d_c(x_i, x_j)$ is minimal. Switching the role of i and j if necessary, we can assume $j = (i+t)_b$. Denote B_i by u and B_{j+1} by v. Then the fragment between two down spins σ_u (the left-most one) and σ_v (the right-most one) in the configuration σ is:

$$-\boxed{x_i}-\boxed{x_{i+1}}-\boxed{x_{i+2}}\cdots-\boxed{x_{i+t-1}}-\boxed{x_j}- \ . \tag{3.12}$$

Without loss of generality, we may assume $x_j > x_i + 1$. Since $d_c(x_i, x_j)$ is minimal, we can assert that

$$x_{i+1} = x_{i+2} = \cdots = x_{i+t-1} = x_i + 1.$$

Thus the fragment in (3.12) can be simplified as the following one, which contains $t-1$ blocks $\boxed{x_i + 1}$ in the middle:

$$-\boxed{x_i}-\boxed{x_i+1}-\boxed{x_i+1}\cdots-\boxed{x_i+1}-\boxed{x_j}- \ . \tag{3.13}$$

By applying the switch operator once, we can obtain a new configuration ϕ, which has the same spin structure as that of σ except replacing (3.13) with

$$-\boxed{x_i+1}-\boxed{x_i+1}-\boxed{x_i+1}-\cdots-\boxed{x_i+1}-\boxed{x_j-1}- \ , \tag{3.14}$$

where the number of $\boxed{x_i + 1}$ in the above fragment is t. Now we conclude that:

(1): $I_s(\phi) = I_s(\sigma)$ for $1 \leq s < x_i + 1$;

(2): $I_{x_i+1}(\sigma) = I_{x_i+1}(\phi) + 1$. (Note that $\sigma_u = \sigma_{u+x_i+1} = -1$ and $\phi_{u+x_i+1} = +1$).

This implies that $H(\sigma) > H(\phi)$ by Corollary 3.1.5. $\square$

For any non-regular configuration σ in $\mathrm{CONF}(a, b)$ with $\rho(\sigma) \geq 1$, we know that

$$x_i \in \left\{ \left\lfloor \frac{a}{b} \right\rfloor, \left\lceil \frac{a}{b} \right\rceil \right\}$$

for $0 \leq i \leq b - 1$. Note that if b divides a, then a configuration σ is 1-regular if and only if it is regular. Therefore we may assume $\lceil a/b \rceil = \lfloor a/b \rfloor + 1$ in the remainder of this subsection. To simplify notation, we denote $\lfloor a/b \rfloor$ and $\lceil a/b \rceil$ respectively by $\perp$ and $\top$. Note that $\perp \geq 1$ from the assumption that $a \geq b$. Furthermore, for any $x \in \{\top, \perp\}$, let $\overline{x}$ be the unique element in $\{\top, \perp\}$ that is different from x.

Now we are proceeding to the proof of the second case.

Lemma 3.3.4. *Suppose that σ is a non-regular configuration in* $\mathrm{CONF}(a, b)$ *with $\rho(\sigma) = 1$. Then there exists a configuration ϕ such that $(\sigma, \phi) \in E(\mathcal{G}_{a,b})$ and $H(\sigma) > H(\phi)$.*

Proof. From Lemma 2.1.1 and the assumptions, there exist i and j such that

$$|(x_i + x_{i+1}) - (x_j + x_{j+1})| \geq 2.$$

Consider one pair (x_i, x_j) such that $d_c(x_i, x_j)$ is minimal over all pairs satisfying the above inequality. Without loss of generality, we may also assume that $B_i < B_j$.

Since $\rho(\sigma) = 1$, we have $|x_i - x_j| \leq 1$. In fact, the equality must hold. Otherwise $|x_{i+1} - x_{j+1}| \geq 2$, a contradiction to the 1-regularity of σ. Therefore we need to consider the following two possible cases:

(1): $x_i = x_{i+1} = \perp$ and $x_j = x_{j+1} = \top$;

(2): $x_i = x_{i+1} = \top$ and $x_j = x_{j+1} = \perp$.

Here we will prove the lemma for Case (2), which can be easily modified for Case (1) as well. With this additional assumption, we claim that the following fact holds.

Fact: There are exactly t copies of the block $\boxed{\perp - \top -}$ between the block $\boxed{x_i - x_{i+1} -}$ and $\boxed{x_j - x_{j+1} -}$ in σ for some $t \in [0, \lfloor a/4 \rfloor]$.

This fact can be verified as follows. If $j = i + 2$, then $t = 0$. Otherwise the components of up spins between x_i and x_{j+1} can be expressed schematically in the following way:

$$\top\top a_1 a_2 \cdots a_s \perp\perp,$$

where $s \geq 1$ and $a_t \in \{\perp, \top\}$ for $1 \leq t \leq s$. From the minimality of $d_c(x_i, x_j)$, we deduce that

$$a_1 = \perp, \quad a_s = \top \ \text{ and } \ a_t \neq a_{t+1} \ \text{ for } \ 1 \leq t < s,$$

which implies that $s = 2t$ and $a_1 a_2 \cdots a_s$ is formed by t blocks of $\boxed{\perp\top}$. This completes the verification of the fact by noticing that there exists one down spin between each component of up spins.

By the above fact, σ contains the following fragment of spins:

$$-\boxed{\top}-\boxed{\top}-\boxed{\perp}-\boxed{\top}-\cdots-\boxed{\perp}-\boxed{\top}-\boxed{\perp}-\boxed{\perp}- \ .$$

By applying the switch operator S, we obtain the following configuration ϕ, which remains the same as σ except for replacing the above fragment with

$$-\boxed{\top}-\boxed{\perp}-\boxed{\top}-\boxed{\perp}-\cdots-\boxed{\top}-\boxed{\perp}-\boxed{\top}-\boxed{\perp}- \ .$$

Recall that $\top = \lceil a/b \rceil$ and $\perp = \lfloor a/b \rfloor$. Put $L := 2\lfloor a/b \rfloor + 2$. Then we have $I_l(\sigma) = I_l(\phi)$ for $1 \leq l < L$ and $I_L(\sigma) > I_L(\phi)$. Together with Corollary 3.1.5, we conclude that $H(\sigma) > H(\phi)$. $\square$

Now we consider the general case that σ is a non-regular configuration in $\mathrm{CONF}(a, b)$ with $\rho(\sigma) = h - 1$ for some $h \in [3, b]$. As before, we need some structure information about the configurations that are $(h-1)$-regular but not h-regular. We begin with some notation.

Given a vector $A = \langle a_i, a_{i+1}, \cdots, a_j\rangle$, the *contraposition* of X, is defined as

$$A^T := \langle a_j, a_{j-1}, \cdots, a_{i+1}, a_i\rangle.$$

When this vector is a segment of the characteristic sequence of the configuration σ, it will also be written as $\underline{a_i a_{i+1} \cdots a_j}$, and is called a *h-block* (or a *block* for simplicity) of σ. Here we will use $A|B$ to denote two adjacent blocks.

Now we state the following stability lemma, whose proof will be presented in Subsection 3.3.4.

Lemma 3.3.5 (stability lemma)**.** *Suppose that σ is a configuration in $\mathrm{CONF}(a, b)$ that is $(h-1)$-regular but not h-regular for some $h \in [3, b]$. Then there exists a block $X := \underline{x a_1 \ \cdots \ a_{h-2} x}$ such that*

(1): $X^T = X$, i.e., $\langle a_1, \cdots, a_{h-2}\rangle^T = \langle a_1, \cdots, a_{h-2}\rangle$;

(2): If we put $Y := \underline{\overline{x} a_1 \ \cdots \ a_{h-2} \overline{x}}$ and $Z := \underline{\overline{x} a_1 \ \cdots \ a_{h-2} x}$, then the characteristic sequence of σ contains

$$X|Z_0|Z_1|\cdots|Z_{k-1}|Y \ ,$$

where X and Y could be adjacent, i.e., $k = 0$.

Note that the above lemma can be regarded as a generalization of the fact used in the proof of Lemma 3.3.4.

Lemma 3.3.6. *Suppose that σ is a non-regular configuration in* $\mathrm{CONF}(a, b)$ *with $\rho(\sigma) = h - 1$ for $3 \leq h \leq b$. Then there exists a configuration ϕ such that $(\sigma, \phi) \in E(\mathcal{G}_{a,b})$ and $H(\sigma) > H(\phi)$.*

Proof. Without loss of generality, the characteristic sequence of σ contains the fragment

$$\top a_1 \ \cdots \ a_{h-2}\top|\bot a_1 \cdots a_{h-2}\top| \ \cdots \ |\bot a_1 \cdots a_{h-2}\top|\bot a_1 \cdots a_{h-2}\bot \tag{3.15}$$

from Lemma 3.3.5. Here the left-most (resp. right-most) block is X (resp. Y), and the middle blocks are Z.

Now we can obtain a configuration $\phi = \mathcal{S}(\sigma)$ whose characteristic sequence is the same as that of σ except replacing the above fragment by

$$\top a_1 \ \cdots \ a_{h-2}\bot|Z_{r-1}^T|Z_{r-2}^T|\cdots|Z_1^T|Z_0^T|\top a_1 \cdots a_{h-2}\bot \ ,$$

which is equivalent to

$$\top a_1 \ \cdots \ a_{h-2}\bot|\top a_1 \cdots a_{h-2}\bot| \ \cdots \ |\top a_1 \cdots a_{h-2}\bot|\top a_1 \cdots a_{h-2}\bot \ , \tag{3.16}$$

since $\langle a_1, \cdots, a_{h-2}\rangle^T = \langle a_1, \cdots, a_{h-2}\rangle$.

Now we claim that the Y block is the minimal h-block in σ, in the sense that for any h-block $Y' := \underline{x_0, \cdots, x_{h-1}}$ in the characteristic sequence in σ, we have

$$\bot + a_1 + \cdots + a_{h-2} + \bot \leq x_0 + \cdots + x_{h-1} \ .$$

If this fails, then we can obtain a contradiction to the fact that σ is $(h-1)$-regular by comparing the block Y' with X. Furthermore, ϕ also does not contain any h-block that is smaller than Y'.

Putting

$$L := \bot + a_1 + \cdots + a_{h-2} + \bot + h,$$

then we can assert that

(1): $\vartheta_j^i(\sigma) = \vartheta_j^i(\phi)$ if $i < h$; (Note that $\Xi_i(\sigma) = \Xi_i(\phi)$ for $i < h$, since σ and ϕ are both $(h-1)$-regular.)

(2): $\vartheta_L^h(\sigma) > \vartheta_L^h(\phi)$;(This is clear from the construction.)

(3): $\vartheta_j^h(\sigma) = \vartheta_j^h(\phi) = 0$ if $j < L$;(This is because any h-block in σ or ϕ is equal to or bigger than Y.)

(4): $\vartheta_j^i(\sigma) = \vartheta_j^i(\phi) = 0$ if $i > h$ and $j \leq L$. (This is because any $(h+1)$-block in σ or ϕ is strictly bigger than Y.)

Together Lemma 3.1.3, the above facts imply that

$$I_L(\sigma) > I_L(\phi) \text{ and } I_l(\sigma) = I_l(\phi) \text{ for } l < L,$$

which completes the proof of the lemma via Corollary 3.1.5. $\square$

3.3.4 The stability lemma

We end Chapter 3 with this subsection, which is devoted to proving the stability lemma.

The Proof of Lemma 3.3.5: Suppose that σ is a configuration CONF(a, b) with $\rho(\sigma) = h - 1$ for $h \in [3, b]$. From Lemma 2.1.1, we assert that there exists a pair (i, j) such that

$$|(x_i + x_{i+1} + \cdots + x_{i+h-1}) - (x_j + x_{j+1} + \cdots + x_{j+h-1})| \geq 2. \quad (3.17)$$

Furthermore, let (i, j) be a pair of the indices such that $d_c(x_i, x_j)$ is minimal over all pairs satisfying (3.17).

Clearly (3.17) leads to two cases to consider. Here we will prove the lemma for the following case (i.e., $x = \top$) while the other case is similar:

$$(x_i + x_{i+1} + \cdots + x_{i+h-1}) - (x_j + x_{j+1} + \cdots + x_{j+h-1}) \geq 2, \quad (3.18)$$

Denoting the block $\underline{x_i x_{i+1} \cdots x_{i+h-1}}$ by X and $\underline{x_j x_{j+1} \cdots x_{j+h-1}}$ by Y, Now we claim that the block X and Y satisfies the requirement in Lemma 3.3.5:

Claim I: In (3.18), we have $x_i = x_{i+h-1} = \top$, $x_j = x_{j+h-1} = \perp$ and $x_{i+t} = x_{j+t}$ for $1 \leq t \leq h-2$.

Proof. If $x_i \le x_j$, then (3.18) implies that

$$(x_{i+1} + \cdots + x_{i+h-1}) - (x_{j+1} + \cdots + x_{j+h-1}) \ge 2,$$

a contradiction to the fact that ϕ is $(h-1)$-regular. Similarly, we can show that $x_{i+h-1} > x_{j+h-1}$ and hence complete the proof of the first part.

If the conclusion of the second part fails, then $x_{i+t} \ne x_{j+t}$ for some $t \in [1, h-2]$. Consider the minimal t such that $x_{i+p} = x_{j+p}$ for all $p < t$. Now if $x_{i+t} > x_{j+t}$, then we have

$$(x_{i+1} + \cdots + x_{i+t}) - (x_{j+1} + \cdots + x_{j+t}) \ge 2.$$

On the other hand, if $x_{i+t} < x_{j+t}$, then we can conclude that

$$(x_{i+t+1} + \cdots + x_{i+h-1}) - (x_{j+t+1} + \cdots + x_{j+h-1}) \ge 2.$$

In both cases, we obtain a contradiction to the fact that σ is $(h-1)$-regular, which completes the proof of the second part. $\square$

From Claim I, the X block can be written as $\underline{\top a_1 \cdots a_{h-2} \top}$ while the Y block as $\underline{\bot a_1 \cdots a_{h-2} \bot}$ with $a_s \in \{\bot, \top\}$ $(1 \le s \le h-2)$. Furthermore, we have the following claim:

Claim II: With the notation above, $\langle a_1 \cdots a_{h-2} \rangle = \langle a_1 \cdots a_{h-2} \rangle^T$.

Proof. If $a_1 \ne a_{h-2}$, then $a_1 > a_{h-2}$ or $a_1 < a_{h-2}$. In the first case, we have

$$(\top + a_1) - (a_{h-2} + \bot) \ge 2$$

by considering the left fragment of the X block and the right fragment of the Y block.

On the other hand, in the second case we know

$$(a_{h-2} + \top) - (\bot + a_1) \ge 2$$

by considering the right fragment of the X block and the left fragment of the Y block.

Since in both cases we obtain a contradiction to the fact that σ is 2-regular, we conclude that $a_1 = a_{h-2}$. Similarly, we can prove that $a_t = a_{h-1-t}$ for $1 \le t \le h-2$. $\square$

Let Z be the block formed by $\underline{\perp a_1 a_2 \cdots a_{h-2} \top}$. Together with Claim I and II, the following claim completes the proof of the lemma:

Claim III: With the notation above, there are exactly k ($k \in \mathbb{N}$) copies of the Z block between the block X and Y in the characteristic sequence of the configuration σ.

Proof. Without loss of generality, let $j = i + h + t \pmod b$ for $t = d_c(x_i, x_j)$. In other words, we have the following fragment between the X block (left) and the Y block (right) in the characteristic sequence of σ:

$$\underline{\top a_1 \cdots a_{h-2} \top}\ c_0 \cdots c_{t-1}\ \underline{\perp a_1 \cdots a_{h-2} \perp}\ . \tag{3.19}$$

Clearly the following two facts imply Claim III:

Fact 1: h divides t. That is, $hk = t$ holds for some k;
Fact 2: $c_0 \cdots c_{t-1}$ consists of k copies of the Z block.

If **Fact 1** fails, then $t = kh + p$ holds for some integers k and p with $1 \le p \le h-1$. Therefore we can divide $c_0 \cdots c_{t-1}$ into the following groups:

$$c_0^0, c_1^0 \cdots, c_{h-1}^0 \mid c_0^1, c_1^1 \cdots, c_{h-1}^1 \mid \cdots \mid c_0^{k-1}, c_1^{k-1} \cdots, c_{h-1}^{k-1} \mid d_0, \cdots, d_{p-1}.$$

Here $c^q = \{c_0^q, c_1^q \cdots, c_{h-1}^q\}$ denotes the q-th group in the decomposition and the low index in c_s^q denotes the relative position of c_s^q in the group c^q.

Now we claim the following fact holds:

Fact 3: For any $0 \le q \le k-1$, $c_0^q, c_1^q \cdots, c_{h-1}^q$ is a copy of the Z block.

The above fact can be verified by induction on q. For the base case $q = 0$, consider the following left-most segment in (3.19):

$$\top a_1 a_2 \cdots a_{h-2} \top \mid c_0^0, c_1^0 \cdots, c_{h-1}^0 \ . \tag{3.20}$$

Then it suffices to show that $c_0^0 = \perp$, $c_{h-1}^0 = \top$ and $c_s^0 = a_s$ for $1 \leq s \leq h-2$ by the following four steps.

Step 1: Firstly we prove $c_0^0 = \perp$. Conversely, if $c_0^0 = \top$, then the X block in (3.18) can be replaced by $a_1 a_2 \cdots a_{h-2} \top | c_0^0$, a contradiction to the minimality of $d_c(x_i, x_j)$.

Step 2: Secondly we prove $c_1^0 = a_1$. Note that $c_1^0 \geq a_1$ because otherwise $(\top + a_1) - (\perp + c_1^0) \geq 2$, a contradiction to the fact that σ is 2-regular. On the other hand, we also have $c_1^0 \leq a_1$. If not, then $c_1^0 = \top$ and $a_1 = \perp$, which implies $\perp + c_1^0 = \top + a_1$. Therefore we can replace the block X in (3.18) by $a_2 \cdots a_{h-2} \top | c_0^0, c_1^0$, a contradiction to the minimality of $d_c(x_i, x_j)$.

Step 3: This is the induction step. For $1 < s < h-1$, we need to prove that $c_s^0 = a_s$ with the assumption that $c_0^0 = \perp$ and $c_l^0 = a_l$ for $1 \leq l \leq s-1$. In other words, (3.20) can be reformulated as the following one:

$$\top a_1 a_2 \cdots a_{s-1} a_s \cdots a_{h-2} \top \mid \perp a_1 \cdots a_{s-1} c_s^0 \cdots c_{h-1}^0 \ .$$

Now if $c_s^0 \neq a_s$, we need to consider the following two cases:

Case i): $c_s^0 > a_s$. Then $c_s^0 = \top$ and $a_s = \perp$. From $c_s^0 + \perp = \top + a_s$, we have

$$\top + a_1 + \cdots + a_{h-2} + \top = a_{s+1} + \cdots + \top + \perp + \cdots + a_{s-1} + c_s^0.$$

Therefore the block X in (3.18) can be replaced by $a_{s+1} \cdots \top \perp \cdots a_{s-1} c_s^0$, a contradiction to the minimality of $d_c(x_i, x_j)$.

Case ii): $c_s^0 < a_s$. In this case we have

$$(\top + a_1 + a_2 + \cdots + a_{s-1} + a_s) - (\perp + a_1 + \cdots + a_{s-1} + c_s^0) \geq 2 \ ,$$

a contradiction to the fact that σ is $(h-1)$-regular.

Step 4: The last step is to prove that $c^0_{h-1} = \top$. If not, then we can replace Y in (3.18) by c^0, a contradiction to the minimality of $d_c(x_i, x_j)$.

From Step 1-4, we complete the proof of **Fact 3** for the base case $q = 0$, i.e., $c^0 = Z$ (this means that c^0 is a copy of the Z block). Now we proceed to the induction step: if $c^s = Z$ for $0 \leq s \leq q-1$, then

$$c^q = Z \quad \text{ for } \; q \in [1, k-1].$$

From the assumption, we have the following fragment:

$$\top a_1 a_2 \cdots a_{h-2} \top \mid \cdots \mid \perp a_1 a_2 \cdots a_{h-2} \top \mid c^q_0, c^q_1 \cdots, c^q_{h-1} \quad .$$

Now we need to prove that $c^q_0 = \perp$, $c^q_{h-1} = \top$ and $c^q_u = a_u$ for $1 \leq u \leq h-2$. This can be done by an argument similar to the above four steps. The details are omitted to save space.

Therefore we complete the proof of **Fact 3**. To sum up, now we have the following right segment in the fragment (3.19):

$$a_0 \; a_1 \; a_2 \; \cdots \; a_{h-2} \; \top \mid d_0 \cdots d_{p-1} \mid \perp \; a_1 \; a_2 \; \cdots \; a_{h-2} \; \perp, \tag{3.21}$$

where $a_0 = \top$ if $k = 0$ and $a_0 = \perp$ otherwise.

Now we claim that $d_0 = \perp$. If this fails, then the X block can be replaced by $a_1 \; a_2 \; \cdots \; a_{h-1} \; \top \mid d_0$. Furthermore, we also have $d_{p-1} = \top$. If not, then we can obtain a contradiction via replacing the Y block by $\perp \mid \perp \; a_1 \; a_2 \; \cdots \; a_{h-1}$. Thus $k \geq 2$ if $k \neq 0$. By a similar argument to Step 2 in proving $c^0 = z$, we can show that $d_s = a_s$ for $1 \leq s \leq p-2$. Therefore the fragment in (3.21) is equivalent to

$$a_0 \; a_1 \; a_2 \; \cdots \; a_{h-2} \; \top \mid \perp a_1 \cdots a_{p-2} \top \mid \perp \; a_1 \; a_2 \; \cdots \; a_{h-2} \; \perp.$$

From the minimality of $d_c(x_i, x_j)$, we conclude that

$$\top + a_1 + \cdots + a_{p-1} > \bot + a_1 + \cdots + a_{p-2} + \top > a_{h-p+1} + \cdots + a_{h-2} + \bot. \quad (3.22)$$

Here the first inequality holds because otherwise the X block can be replaced by the fragment

$$a_p \cdots a_{h-2} \top | \bot a_1 \cdots a_{p-2} \top \ .$$

Similarly, the second inequality holds, otherwise the block Y can be replaced by the fragment

$$\bot a_1 \cdots a_{p-2} \top | \bot \ a_1 \ \cdots a_{h-p} \quad .$$

On the other hand, from Claim II, we know $a_s = a_{h-1-s}$ for $1 \leq s \leq h-2$. Together with (3.22), it implies that

$$\top + a_{p-1} > \bot + \top > a_{p-1} + \bot. \quad (3.23)$$

From which we can assert that $\bot < a_{p-1} < \top$, a contradiction to the fact that $a_{p-1} \in \{\bot, \top\}$. Therefore we have $p = 0$, which completes our proof of **Fact 1**: $t = kh$ for some k. Furthermore, this proof also implies the correctness of **Fact 2**. Therefore we have completed the proof of Claim III. □

Since Lemma 3.3.5 follows directly from Claim I, II and III, we also complete the proof of the lemma as well. □

Note

Let us remark here that the links between ground states in the Kawasaki-Ising model and maximally even configurations was firstly investigated in [22, 23]. Independently, we studied the connections between ground states and regular configurations in [15, 14]. After publishing [14], we realized the equivalence between maximally even and regularity, which leads to a new proof of Theorem 3.2.1.

Chapter 4

Cycle Packing

In this chapter we study the cycle packing problem for shift digraphs, the Cayley digraphs of $\mathbb{Z}_n$ with two generators. That is, we show that the maximal number of vertex-disjoint cycles in shift digraphs is determined by its size and girth. In addition, we can find a shortest cycle such that it produces enough disjoint copies by rotating.

4.1 Cycles

In this section, we propose a scheme to encode the cycles in shift digraphs via the configurations in the Kawasaki-Ising model. This scheme makes it possible to apply the theory of regular configurations to solve the cycle packing problem of shift digraphs.

Given a cycle $C = (v_0, v_1, \cdots, v_{d-1})$ in $\mathrm{Cay}(\mathbb{Z}_n, \{l, m\})$, its *difference sequence* is defined as

$$\nabla(C) := ((v_1 - v_0)_n, \cdots, (v_{d-1} - v_{d-2})_n, (v_0 - v_{d-1})_n).$$

Note that the sequence $\nabla(C)$ consists only of two numbers, l and m. In other words, it is a word over the binary alphabet $\{l, m\}$. Denoting the number of the occurrences of l (resp. m) in this sequence by b (resp. a), $\nabla(C)$ can be regarded as a configuration in $\mathrm{CONF}(a, b)$, where $a+b = d$.

More precisely, here l represents a down spin and m represents an up spin.

This configuration is also denoted by σ_C. By this scheme, any cycle in a shift digraph can be encoded as a pair (v_0, σ), where v_0 is the starting vertex and σ is the *coding configuration*. Note that the same cycle can be encoded as two different pairs, say (v, σ) and (u, σ'), by choosing two different starting vertices, but they satisfy the relation $\sigma \sim \sigma'$, i.e., they are equivalent up to the shift operator.

Example 5. An cycle in $\mathrm{Cay}(\mathbb{Z}_9, \{1, 3\})$.

Consider the cycle $C = (0, 1, 4, 5, 6)$ in $\mathrm{Cay}(\mathbb{Z}_9, \{1, 3\})$. Clearly we have $\nabla(A_C) = (1, 3, 1, 1, 3)$ and C can be encoded as the pair $(0, \sigma)$, where $\sigma = (B, R, B, B, R)$ is a configuration in $\mathrm{CONF}(2, 3)$.

On the other hand, given any pair (v_0, σ), it is not difficult to find the cycle corresponding to it in $\Gamma = \mathrm{Cay}(\mathbb{Z}_n, \{l, m\})$ if such cycle is contained in Γ. Note that there exists a cycle C in $\mathrm{Cay}(\mathbb{Z}_n, \{l, m\})$ such that $C = C_{0,\sigma}$ for some $\sigma \in \mathrm{CONF}(a, b)$ if and only if $n \mid am + bl$.

When the configuration σ is clear from the context, $C_{v,\sigma}$ will also be simply written as $\underline{v}$. In this setting, the cycle $\underline{0}$, which plays an important role in the following analysis, is called the *generic cycle* of σ and its vertex set is denoted by V_σ.

Definition 4.1.1. *Given a set $B \subseteq V(\mathrm{Cay}(n; \{l, m\}))$, its* difference set *$D(B)$ is defined to be $\{(b_i - b_j)_n \mid \forall\ b_i, b_j \in B\}$.*

From the above definition, we have the following direct consequences.

Proposition 4.1.1. *Given a non-empty set $B \subseteq V(\mathrm{Cay}(n; \{l, m\}))$, the following two facts hold:*

(1): $0 \in D(B)$ and $n \notin D(B)$;

(2): $x \in D(B)$ implies $(-x)_n \in D(B)$.

□

In the remainder of this section, we fix a coding configuration σ for some cycles in a shift digraph $\Gamma = \text{Cay}(n, \{l, m\})$ and study the properties of $D(V_\sigma)$. Here we assume $\sigma = (\iota_0, \cdots, \iota_{a+b-1})$, where $\iota_i \in \{l, m\}$ for each i, belongs to $\text{CONF}(a, b)$ and its characteristic sequence is $(x_0, \cdots, x_{b-1})$. As before, we will also denote $a + b$ by d.

Proposition 4.1.2. *For any two cycles $\underline{s}$ and $\underline{t}$ in $\text{Cay}(n, \{l, m\})$ that are encoded by σ, $\underline{s} \cap \underline{t} \neq \emptyset$ if and only if $(s - t)_n \in D(V_\sigma)$.*

Proof. Clearly we have $V_\sigma = (0, \kappa_0, \kappa_1, \cdots, \kappa_{d-2})$, where

$$\kappa_i = (\sum_{j=0}^{i} \iota_j)_n$$

for $0 \leq i \leq d - 2$. Furthermore we have

$$V(\underline{t}) = (t, (t + \kappa_0)_n, (t + \kappa_1)_n, \cdots, (t + \kappa_{d-2})_n)$$

and

$$V(\underline{s}) = (s, (s + \kappa_0)_n, (s + \kappa_1)_n, \cdots, (s + \kappa_{d-2})_n).$$

Thus $\underline{s} \cap \underline{t} \neq \emptyset$ if and only if there exists a pair of indices i, j such that

$$s + \kappa_i \equiv t + \kappa_j \pmod{n},$$

which is equivalent to $(s - t)_n \in D(V_\sigma)$ from Proposition 4.1.1. □

The above proposition implies the following corollary, whose proof is straightforward.

Corollary 4.1.3. *For any two cycles $\underline{s}$ and $\underline{t}$ in $\text{Cay}(n, \{l, m\})$ that have the same encoding configuration, $\underline{s} \cap \underline{t} \neq \emptyset$ if and only if $\underline{s+1} \cap \underline{t+1} \neq \emptyset$.*
□

The difference set of V_σ can be characterized by the following proposition.

Proposition 4.1.4. *Given a configuration* $\sigma = (\iota_0, \cdots, \iota_{d-1})$, *the difference set of its generic cycle is:*

$$D(V_\sigma) = \{\iota_i + \iota_{i+1} + \cdots + \iota_{i+s} \mid 0 \leqslant i \leqslant d-1, 0 \leqslant s < d-1\} \cup \{0\}.$$

Proof. $\forall x, y \in V_\sigma$, if $x = y$, then $x - y = 0$; otherwise we have: $x = \iota_0 + \iota_1 + \cdots + \iota_p$ and $y = \iota_0 + \iota_1 + \cdots + \iota_q$ for two distinct numbers p and q in $[1, d-1]$. If $p > q$, then

$$x - y = \iota_{p+1} + \cdots + \iota_q.$$

Otherwise from the fact that$(\iota_0 + \iota_1 + \cdots + \iota_{d-1}) = n$, we have

$$\begin{aligned}
(x-y)_n &= x + n - y \\
&= (\iota_0 + \iota_1 + \cdots + \iota_p) + (\iota_0 + \iota_1 + \cdots + \iota_{d-1}) - (\iota_0 + \iota_1 + \cdots + \iota_q) \\
&= (\iota_0 + \iota_1 + \cdots + \iota_p) + (\iota_{q+1} + \cdots + \iota_{d-1}) \\
&= \iota_{q+1} + \cdots + \iota_{d-1} + \iota_0 + \iota_1 + \cdots + \iota_p\ .
\end{aligned}$$

□

Given a configuration σ and an integer $j \in [1, b]$, μ_j and ξ_j were defined in Chapter 2 (see Definition 2.1.2) as

$$\mu_j = \min_{0 \leqslant i \leqslant b-1} \{x_i + x_{i+1} + \cdots + x_{i+j-1}\},$$

and

$$\xi_j = \max_{0 \leqslant i \leqslant b-1} \{x_i + x_{i+1} + \cdots + x_{i+j-1}\}.$$

Here $(x_0, x_1, \cdots, x_{b-1})$ is the characteristic sequence of σ and the subscripts in x_i are calculated modulo b. Furthermore, we use the convention that $\mu_{-1} = \mu_0 = 0$ and $\xi_{b+1} = a - 1$. Then we have the following proposition.

Proposition 4.1.5. *Given a configuration* σ *in* $\mathrm{CONF}(a, b)$, *then we* *have*

$$D(V_\sigma) = \{p_j m + jl \mid 0 \leqslant j \leqslant b,\ \mu_{j-1} \leqslant p_j \leqslant \xi_{j+1}\}.$$

Proof. The boundary cases can be verified directly and the other cases follow from Proposition 4.1.4 by considering the number of l's in the expressions of the elements in $D(V_\sigma)$. □

To illustrate the concepts mentioned so far, we consider the following example.

Example 6. A cycle in $\Gamma = \mathrm{Cay}(11, \{1, 3\})$.

Consider the cycle $C = (0, 3, 4, 7, 10)$ in Γ. Here the parameters of the shift digraph are $n = 11, l = 1, m = 3$. Then $\nabla(C) = (3, 1, 3, 3, 1)$ and C can be encoded as the pair $(0, \sigma)$ with $\sigma = (3, 1, 3, 3, 1) \in \mathrm{CONF}(3, 2)$. Since 3 denotes the up spin and 1 denotes the down spin, σ can be also written as (R, B, R, R, B). Clearly, for this cycle we have $a = 3, b = 2$ and $\delta(C) = 1$. Furthermore, the characteristic sequence for σ is $(2, 1)$. Therefore, by definition we have

$$\mu_{-1} = \mu_0 = 0, \quad \mu_1 = 1, \quad \mu_2 = 3,$$

and

$$\xi_1 = 2, \quad \xi_2 = 3, \quad \xi_3 = 2.$$

To sum up, we have the following table.

j	μ_{j-1}	ξ_{j+1}	p_j	$j + 3p_j$
0	0	2	{0,1,2}	{0,3,6}
1	0	3	{0,1,2,3}	{1,4,7,10}
2	1	2	{1,2}	{5,8}

Then we can verify directly that the union of the last two columns gives us exactly the same set as

$$D(V_\sigma) = D(C) = \{0, 1, 3, 4, 5, 6, 7, 8, 10\}.$$

4.2 Regularity and disjointness

In this section we investigate one connection between the regularity and the disjointness of cycles in shift digraphs.

Note that $\mathrm{Cay}(n; \{l, m\})$ contains a cycle consisting of a type II arcs (that are generated by l) and b type I arcs (that are generated by m) if and only if $n \mid am + bl$.

Theorem 4.2.1. *Given a digraph* $\mathrm{Cay}(n; \{l, m\})$ *and a pair integers* (a, b) *such that* $n \mid am + bl$*, suppose that* σ *is the regular configuration in* $\mathrm{CONF}(a, b)$ *and denote* $\lfloor n/(a + b)\rfloor$ *by* k*. Then the following set:*

$$\mathcal{C} = \{\underline{0}, \underline{\beta(m - l)}, \cdots, \underline{\beta(k - h)(m - l)}\},$$

where $\underline{i} = C_{i,\sigma}$*,* $h = \gcd(l, m)$ *and* $\beta = 1/h$*, is a collection of pairwise disjoint cycles in* $\mathrm{Cay}(n; \{l, m\})$*.*

Proof. From the assumption, we have $am + bl = tn$ for some $t \in \mathbb{N}^+$. By contradiction, if the theorem fails, then we have

$$q\beta(m - l) \in D(V_\sigma)$$

from Proposition 4.1.2. Together with Proposition 4.1.5, this implies the following equation has an integer solution (j, q) such that $0 \leqslant j \leqslant b$ and $1 \leqslant q \leqslant k - h$:

$$q\beta(m - l) \equiv p_j m + jl \pmod n. \tag{4.1}$$

Let

$$r := \left\lfloor \frac{q\beta(m - l)}{n} \right\rfloor = \left\lfloor \frac{tq\beta(m - l)}{am + bl} \right\rfloor.$$

Since $p_j m + jl < n$ from the definition, equation (4.1) can be simplified as

$$q\beta(m - l) = p_j m + jl + \frac{r}{t}(am + bl), \tag{4.2}$$

which gives us

$$m = l\frac{tq\beta + rb + tj}{tq\beta - tp_j - ra}. \tag{4.3}$$

Since $\gcd(l, m) = 1/\beta$, equation (4.3) has integer solutions if and only if the following two equations have integer solutions for some $s \geq 1$:

$$tq\beta + rb + tj = sm\beta, \tag{4.4}$$

$$tq\beta - tp_j - ra = sl\beta. \tag{4.5}$$

By eliminating q from the above two equations we obtain

$$s(m - l)\beta = r(a + b) + t(j + p_j), \tag{4.6}$$

which yields

$$m = l + \frac{r(a + b) + t(j + p_j)}{s\beta}. \tag{4.7}$$

On the other hand, equation (4.4) implies

$$q = \frac{sm\beta - tj - rb}{t\beta}. \tag{4.8}$$

Since

$$k = \left\lfloor \frac{am + bl}{t(a + b)} \right\rfloor,$$

we have

$$t(a + b)k \leq am + bl. \tag{4.9}$$

Together with $0 \leq q \leq k - h$, it implies

$$t(a + b)(q + h) \leq am + bl. \tag{4.10}$$

Substituting (4.8) into the above equation, we can assert that

$$t(a + b)(h + \frac{sm\beta - tj - rb}{t\beta}) \leq am + bl,$$

which can be further simplified as

$$(a + b)(th\beta + sm\beta - tj - rb) \leq am\beta + bl\beta.$$

Since $h\beta = 1$, we deduce from the above equation that

$$am(s-1)\beta + bsm\beta + t(a+b) \leq bl\beta + (tj + rb)(a+b).$$

Using the fact that $s \geq 1$, we further obtain

$$bsm\beta + t(a+b) \leq bl\beta + (tj + rb)(a+b).$$

Substituting (4.7) into the above equation, we see that the following inequality holds:

$$bs\beta(l + \frac{r(a+b) + t(j+p_j)}{s\beta}) + t(a+b) \leq bl\beta + (rb + tj)(a+b),$$

which can be simplified as

$$bl\beta(s-1) + bt(j+p_j) + t(a+b) \leq tj(a+b).$$

Using the fact that $s \geq 1$ again, we obtain

$$bt(j+p_j) + t(a+b) \leq tj(a+b).$$

Since $t > 0$, we conclude that

$$bp_j + a + b \leq aj. \tag{4.11}$$

If $b = 0$, then $j = 0$ since we assume that $j \in [0, b]$. Hence (4.11) implies $a = 0$, a contradiction. Otherwise, we have $b > 0$ and the following inequality holds:

$$1 + p_j \leqslant \frac{a}{b}(j-1).$$

Since $\mu_{j-1} \leqslant p_j$, we have

$$1 + \mu_{j-1} \leqslant \frac{a}{b}(j-1). \tag{4.12}$$

Therefore if the theorem fails, then there must exist some $j \in [0, b]$

such that (4.12) holds. On the other hand, since σ is regular, we know that

$$1 + \mu_{j-1} > \frac{a}{b}(j-1)$$

holds for all $j \in [0, b]$ from Lemma 2.1.2. This contradicts (4.12), and hence completes the proof. $\square$

One important case of the above theorem is that l and m are coprime.

Theorem 4.2.2. *Given a digraph* $\mathrm{Cay}(n; \{l, m\}$ *with* l *and* m *being coprime and a pair integers* (a, b) *such that* $n \mid am + bl$*, suppose that* σ *is the regular configuration in* $\mathrm{CONF}(a, b)$ *and denote* $\lfloor n/(a+b) \rfloor$ *by* k*. Then the following set:*

$$\mathcal{C} = \{\underline{0}, \underline{(m-l)}, \cdots, \underline{(k-1)(m-l)}\},$$

where $\underline{i} = C_{i,\sigma}$*, is a collection of pairwise disjoint cycles in* $\mathrm{Cay}(n; \{l, m\})$*.*

$\square$

4.3 Cycle packing number

In this section, we shall use the results in the previous section to show that the cycle packing number of a shift digraph is determined by its girth.

By the definition of girth, the following lemma clearly holds.

Lemma 4.3.1. *For any digraph* D*, its cycle packing number* $\nu_0(D)$ *and girth* $\omega(D)$ *satisfy* $\omega(D)\nu_0(D) \le |V(D)|$*.* $\square$

Now we can state a restricted version of our main result in this section.

Lemma 4.3.2. *Suppose that* $\gcd(n, l, m) = 1$*. Then*

$$\nu_0(\Gamma) = \left\lfloor \frac{n}{\omega(\Gamma)} \right\rfloor$$

holds for the digraph $\Gamma = \mathrm{Cay}(n, \{l, m\})$*.*

Proof. From Lemma 4.3.1, it suffices to show that

$$\nu_0(\Gamma) \geq \left\lfloor \frac{n}{\omega(\Gamma)} \right\rfloor, \tag{4.13}$$

which can be proved by considering the following two possible cases.

Case I: $\gcd(l, m) = 1$. From Theorem 4.2.2, any cycle of length d in Γ implies $\nu_0(\Gamma) \geq \lfloor n/d \rfloor$. Now (4.13) follows from the fact that Γ always contains a cycle of length ω.

Case II: $\gcd(l, m) > 1$. Denoting $\gcd(l, m)$ by α, then we know that $\gcd(\alpha, n) = 1$ from the assumption that $\gcd(n, l, m) = 1$. In other words, we have $\alpha \in \mathbb{Z}_n^*$, and hence that Γ is isomorphic to

$$\Gamma' := \mathrm{Cay}(n; \{\alpha^{-1}l, \alpha^{-1}m\}).$$

Now $\gcd(\alpha^{-1}l, \alpha^{-1}m) = 1$. Therefore (4.13) holds for Γ', as we show in Case I. Furthermore, (4.13) holds for Γ as well, since $\nu_0(\Gamma) = \nu_0(\Gamma')$ and $\omega(\Gamma) = \omega(\Gamma')$. This completes the proof of Case II. □

Note that if $\gcd(n, l, m) = \beta$, then $\Gamma = \mathrm{Cay}(n; \{l, m\})$ has β connected components with each of them being isomorphic to the Cayley digraph $\Gamma' = \mathrm{Cay}(n/\beta, \{l/\beta, m/\beta\})$. By this observation, the above lemma can be clearly generalized to the following

Theorem 4.3.3. *Suppose that* $\gcd(n, l, m) = \alpha$. *Then*

$$\nu_0(\Gamma) = \alpha \left\lfloor \frac{n}{\alpha\omega(\Gamma)} \right\rfloor$$

holds for the digraph $\Gamma = \mathrm{Cay}(n; \{l, m\})$. □

From the above theorem, the cycle packing problem for shift digraphs, which is to calculate $\nu_0(\Gamma)$, is reduced to calculate the girth of Γ, which can be solved in $O(n^2)$ time.

4.4 Guessing number

In this section, we present a brief exposition of guessing number and use the results in the previous sections to obtain the bounds of this parameter for a family of digraphs.

Guessing Number was introduced by Riis in studying network coding and circuit complexity [49]. Before presenting its formal definition, here we give an informal description by the following 'game'.

Given a digraph D, we can play a guessing game as follows. Each node is randomly assigned a bit from $\{0, 1\}$ and each node knows only the bit assigned to its in-neighbors but not the one for itself. Now the task for each node is to guess the bit assigned to itself.

Here we are interested in the probability that all nodes can simultaneously correctly guess their bits in the above game. Then the guessing number measures the best probability we can achieve over all allowed protocols.

Now we fix some notations used in this section. Recall that a *configuration* on digraph D is a map from its vertex set $V(D)$ to $\mathbb{Z}_2 := \{0, 1\}$. All such configurations on D form a set Ω; the variables that take values in Ω will be denoted by $x, y, \cdots$. Note that the ring structure of $\mathbb{Z}_2$ induces a natural ring structure on Ω as well.

A *protocol* $\mathcal{P}$ on a digraph D is a map between its configurations such that $\mathcal{P}(x)$ is locally defined, i.e., $\mathcal{P}(x)_v = (f_v)(x_{v_1}, \cdots, x_{v_k})$ for any $v \in V$, where $k = |N_-(v)|$ and $x_{v_i} \in N_-(v)$ for each x_{v_i}. Note that we can also associate a ring structure with the set of all protocols for a given graph, where the composition of two protocols is defined to be point-wise.

Let us remark here that not every map acting on Ω is a protocol on D. For instance, the identity map $\mathcal{I}$, which maps each configuration x to itself, is not a protocol for any simple digraphs.

Given a protocol $\mathcal{P} = (f_v)_{v \in V(D)}$, let

$$\mathrm{Fix}(\mathcal{P}) := \{x \in \Omega \mid x_v = f_v(x_{v_1}, \cdots, x_{v_k}) \text{ for all } v\}$$

be the set formed by the *fixed points* of $\mathcal{P}$.

Definition 4.4.1. *Given a digraph D, its guessing number is defined as*

$$g(D) := \max_{\mathcal{P}} g(D, \mathcal{P}),$$

where $\mathcal{P}$ runs over all allowed protocols on D and $g(D, \mathcal{P}) := \log_2(|\,\mathrm{Fix}(\mathcal{P})|)$.

Now we collect the following observations about guessing numbers, whose proofs are straightforward and can be obtained in [58].

Proposition 4.4.1. *For any digraph $D = (V, E)$ with $|V| = n$, the following assertions hold:*

(a) $0 \leq g(D) \leq |V| - 1$;

(b) *If $H = (V', E')$ is a subgraph of D, then $g(H) \leq g(D)$;*

(c) *If D is a directed cycle, we have $g(D) \geq 1$;*

(d) *If D is acyclic, then $g(D) = 0$.*

(e) *If D is the* disjoint union *of two graphs H_1 and H_2, then $g(D) = g(H_1) + g(H_2)$*

Let $\tau(D)$ be the size of the minimal vertex set S in D such that $D - S$ is acyclic. Recall that H is an *induced subgraph* of D if $(u, v) \in E(H)$ holds for any $(u, v) \in E(D)$ with $u, v \in V(H) \subseteq V(D)$. Now we can state one of our main results in this section as follows.

Theorem 4.4.2. *For any digraph D, we have*

$$\nu_0(D) \leq g(D) \leq \tau(D).$$

Proof. To establish the first inequality, let $\{c_1, \cdots, c_k\}$ be a set of vertex disjoint cycles of D with $k = \nu_0(D)$ and consider the disjoint union c of these cycles. Then the first inequality clearly holds since

$$g(D) \geq g(c) = \sum_{i=1}^{k} g(c_i) \geq k = \nu_0(D)$$

holds from Proposition 4.4.1.

Now the second one follows directly from Proposition 4.4.1 and the following

Claim: For any induced subgraph H of D, we have $g(D) \leq g(H) + |V(D) - V(H)|$.

It is sufficient to establish this claim for the special case $V(D) - V(H) = \{0\}$. For any optimal protocol $\mathcal{P}$ on D, let $\mathcal{P}^i$ denote the induced protocol on H by putting $x_0 = i$ for $i = 0, 1$. That is, if $f(x_0, x_1 \cdots) \in \mathcal{P}$, then $f(i, x_1, \cdots) \in \mathcal{P}^i$. Clearly for any $x \in \mathrm{Fix}(\mathcal{P})$, we have $x|_H \in \mathrm{Fix}(\mathcal{P}^i)$ if and only if $x_0 = i$. In other words, we have $|\mathrm{Fix}(\mathcal{P})| \leq |\mathrm{Fix}(\mathcal{P}^0)| + |\mathrm{Fix}(\mathcal{P}^1)|$, from which we conclude that

$$\begin{aligned} g(D) = \log_2 |\mathrm{Fix}(\mathcal{P})| &\leq \log_2(|\mathrm{Fix}(\mathcal{P}^0)| + |\mathrm{Fix}(\mathcal{P}^1)|) \\ &\leq \log_2 2|\mathrm{Fix}(\mathcal{P}')| \\ &= 1 + g(H) \end{aligned}$$

holds as required, where $\mathcal{P}'$ is an optimal protocol on H. $\square$

One direct consequence of the above theorem is $g(C_n) = 1$. Let us also remark here that the results in this section also hold for other alphabet sets with cardinality greater than 2.

Note that $g(D) \leq \max\{l, m\}$ clearly holds for any connected digraph $D = \mathrm{Cay}(n; \{l, m\})$. Therefore, by Lemma 4.3.2, the above theorem implies directly the following

Corollary 4.4.3. *For any connected digraph $D = \mathrm{Cay}(n; \{l, m\})$, we have*

$$\left\lfloor \frac{n}{\omega(D)} \right\rfloor \leq g(D) \leq \max\{l, m\}.$$

□

In particular, we have a more explicit bound on directed double loops, a subfamily of shift graphs.

Corollary 4.4.4. *For a digraph $D = \mathrm{Cay}(n; \{1, m\})$, we have*

$$\left\lfloor \frac{n}{n + (1-m)\lfloor n/m \rfloor} \right\rfloor \leq g(D) \leq m,$$

Proof. Putting $p := \lfloor n/m \rfloor$ and $r := n - pm$, then it is clear that $n = pm + r$ holds and $\mathrm{Cay}(n; \{1, m\})$ contains a cycle of size $p + r$. Together with Corollary 4.4.3, this establishes the corollary. □

Part II

TBR Graphs

Chapter 5

Introduction

Phylogenetic trees, i.e., leaf-labelled trees, are useful to study evolution relationships in biology and other areas of classification. For some problems, from a given data set one needs to construct a tree that is optimal according to a given criteria, such as maximal parsimony or maximal likelihood, to understand the true evolution history.

As the search space is large: there are $(2n-5)!! = 1 \times 3 \times \cdots \times (2n-5)$ unrooted binary phylogenetic trees for n objects (a result that dates back to [52], see also [53]), many tree (re-)construction problems are intractable and hence heuristic algorithms are popular for practical applications. In such algorithms we start with an initial tree, which could be chosen randomly or in an intelligent way, and apply some types of local changes to find a new tree with better scores for the given criteria until we reach a local optimal solution.

Making local changes is often referred to as a tree rearrangement operation. Besides playing an important role in designing algorithms, these operations are also useful in measuring the similarity between two given trees [50]. For both purposes, it is natural to consider the metric induced by the following operation graphs: the vertex set consists of all trees with a given leaf set and two trees are adjacent if and only if they differ by exactly one operation.

The maximal and minimal degree of such graphs are of special interest for studying the performance of heuristic algorithms as they are closely related to the complexity of a local move. Among the three most common operations: NNI, SPR and TBR (see Section 5.1), in the literature, it is well known that both the NNI graphs and the SPR graphs are regular. But for the TBR graphs, this problem is more involved and we refer the reader to [31] for the best known bounds.

To fill in this gap, we present here the first closed-form formula to calculate the degree of the vertices in the TBR graphs, a quantity also referred to as "the size of the TBR unit-neighborhood" (cf. [31]), and show that it is determined by Γ-index (see Chapter 6), a tree index introduced here to measure the shape of trees. By this formula, we obtain the maximal and minimal degree, as well as the average degree, of the TBR graphs.

Among the trees with a given maximum degree, we show that the tree achieves the minimal Γ-index is a "good tree", which has been extensively studied in computer science [39] and also coincides with the extremal tree of several other graphical indices [33, 25, 57]. Note that some authors (e.g. [39]) have used different terminology, referring good trees as "complete trees". The approach presented here is naive and arguably simpler. More interestingly, here we also obtain a structural characterization of good trees, and provide a principle that may be employed to a general solution to the extremal problems for other indices.

With a multivariate contraction method developed in the context of random searching trees [46], we obtain the mean and variance of the size of the TBR unit-neighborhood of a random tree generated by the Yule-Harding model [66, 29], one of the most famous stochastic models used to generate random phylogenetic trees [1, 2].

By a technique related to homoplasy scores [12], we obtain a current best known lower bound on the diameter of the TBR graphs [28]. Finally, we also characterize the extremal trees for the Γ-index among the trees with given degree sequence and apply the semi-regularity principle to other graphical indices.

The remainder of this part is organized as follows. In Section 5.1 we collect some definitions and notation. Chapter 6 is devoted to the Γ-index and related topics, including a new characterization of good trees by the semi-regularity property. In Chapter 7 we study various properties of TBR graphs, such as the maximal and minimal degree of the TBR graphs, the size of the TBR unit-neighborhood of Yule-Harding random trees, and the diameter of the TBR graphs. The results in Chapter 6 are contained in the author's joint work with Hua Wang [61], and Section 7.1 is based on a joint work with Peter Humphries [32].

5.1 Definitions and notation

Some basic definitions and notation for Part II are collected in this section and we refer the readers to [53] for a more detailed exposition of the concepts mentioned here.

All graphs in this part will be finite, simple and undirected. For any vertex v in a graph, let $\deg(v)$ denote the *degree* of v, i.e., the number of edges incident to v. A *tree* $T = (V, E)$ is a connected, acyclic graph. $V(T)$ and $E(T)$ denote the vertex set and edge set of a tree T. We refer to vertices of degree 1 of T as *leaves*, which form the *leaf set* $L(T)$. The edges incident to some leaf are called *pendant edges*, and a *cherry* is a pair of leaves $\{x, y\}$ adjacent to the same interior vertex.

The unique path connecting two vertices v, u in T will be denoted by $P_T(v, u)$. For a tree T and two vertices v, u of T, the *distance* $\mathrm{dist}_T(v, u)$ between them is the number of edges on the path $P_T(v, u)$.

In this part, our main concern is unrooted binary *phylogenetic trees*, that is, bijectively leaf-labelled trees without a specified root in which every interior vertex has degree 3. We denote by $\mathcal{T}_n$ the set of all such trees with the same leaf set $\{1, \ldots, n\}$.

For our purpose, we also need to consider rooted trees. Here we call a tree (T, r) *rooted at the vertex* r (or just T if it is clear what the root is) by specifying a vertex $r \in V(T)$. The *height* of a vertex v of a rooted tree T with root r is $h_T(v) = \mathrm{dist}_T(r, v)$, and the *height* of T is just the greatest height of its vertices. Note that this concept is also referred to as the *depth* in many literatures.

For any two different vertices u, v in a rooted tree (T, r), we say that v is a *successor* of u and u is an *ancestor* of v if $P_T(r, u) \subset P_T(r, v)$. For a vertex v in a rooted tree (T, r), we use $T(v)$ or T_v to denote the subtree rooted at v, induced by v and all its successors.

In this thesis, we are mainly interested in unrooted binary phylogenetic trees, but the families of trees in the following list are also considered:

- $\mathcal{T}_n^*$: the set of all rooted binary phylogenetic trees with n leaves;
- $\mathbf{R}_n^d$: the set of rooted trees with n leaves such that the number of the successors of any vertex is at most d;
- $\mathbf{U}_n^d$: the set of unrooted trees with vertex degrees not exceeding d.

We will put $\mathcal{T} := \cup_{n=1}^{\infty} \mathcal{T}_n$ and similarly we can define $\mathcal{T}^*$, $\mathbf{U}^d$ and $\mathbf{R}^d$.

In the remainder of this section, we give a brief introduction to three tree rearrangement operations that are commonly studied in literature. Following [3], they are presented below from the most restrictive one to the most general one.

- NNI: Any internal edge of a tree $T \in \mathcal{T}_n$ has four subtrees attached to it. A *nearest neighbor interchange* (NNI) occurs when one subtree on one side of an internal edge is swapped with a subtree on the other side of the edge.
- SPR: A *subtree prune and regraft* (SPR) operation on a tree $T \in \mathcal{T}_n$ involves deleting some edge e from T and thereby pruning a subtree t, and then regrafting the subtree by the same cut edge to a new vertex obtained by subdividing a pre-existing edge in $T - t$.

- TBR: A *tree bisection and reconnection* (TBR) on a tree $T \in \mathcal{T}_n$ consists of two steps: deleting some edge e from T to obtain two subtrees, and subsequently inserting an edge in one (in the case that the other one is an isolated labelled vertex) or both subtrees to form a new tree T' that is distinct from T.

Here we will use the convention that in each step the vertices with degree 2 will be contracted, i.e., any vertex of degree 2 will be deleted and the two edges incident to it will be replaced by a single edge. Therefore any tree in $\mathcal{T}_n$ will remain in the same category after any operations mentioned above.

Clearly, every NNI operation is a SPR operation, and each SPR operation is a TBR operation. For each operation $\ominus \in \{\text{NNI}, \text{SPR}, \text{TBR}\}$, we can associate it with a family of graphs $G_\ominus(n) = (V_n, E_n)$ with $V_n = \mathcal{T}_n$ and E_n consisting of all pairs $\{T_1, T_2\}$ such that T_1 and T_2 differ by one $\ominus$ operation. For abbreviation, given any $T \in \mathcal{T}_n$, we let $\deg_\ominus(T)$ stand for the degree of T in the $\ominus$ graph $G_\ominus(n)$. To avoid trivial cases, in this part we will always assume that $n \geq 4$ holds.

Chapter 6

Γ-index

In this chapter, we introduce Γ-index, a tree index that will be used in Chapter 7 to study the degree distribution of TBR graphs. The extremal trees for this index are studied for several families of trees. In addition, we also obtain a structural characterization of good trees.

6.1 Tree index

Recall that a split $A|B$ of a set X is a bipartition of X into two non-empty disjoint subsets and each edge in a tree T induces canonically a $L(T)$-split. Denoting the number of leaves in a tree or a set $A \subseteq V(T)$ by $|T|$ or $|A|$, then we have

Definition 6.1.1. *For any tree T, the* Γ-index *of T is defined as*

$$\Gamma(T) := \sum_{\{u,v\}\subseteq L(T)} \mathrm{dist}_T(u,v) = \sum_{e\in E} |A_e| \cdot |B_e|, \tag{6.1}$$

where $A_e|B_e$ denotes the $L(T)$-split induced by e.

Clearly, the above definition is well defined, i.e., the equality in (6.1) indeed holds for any tree T. In other words, the sum of the distances between all pairs of leaves in T is equal to the sum of the "weight" of all $L(T)$-splits. There two slightly different formulations are both useful in different contexts.

Note that a similar tree index, the *Wiener Index*

$$W(T) := \sum_{\{u,v\} \subseteq V(T)} \mathrm{dist}_T(v, u)$$

of a tree T, was introduced by Harold Wiener [63] and has been one of the most widely used descriptors in quantitative structure activity relationships. Since the majority of the chemical applications of the Wiener index deal with chemical compounds with acyclic molecular graphs, the Wiener index of trees has been extensively studied over the past years, see [21] and the references therein for details.

For (strict) binary trees, i.e., the degree of each node is either 1 or 3, the Γ-index and the Wiener index are strongly correlated. More precisely, we have $W(T) = 4\Gamma(T) - (8n^2 - 18n + 6)$ in this case, which does not hold for general cases. As we will see in Section 7.2, the Γ-index can also be defined for rooted trees, which is closely related to that of unrooted trees.

6.2 Good trees

In this section, we investigate good trees, a family of trees that has been intensively studied in many areas. For the extremal problem for graphical indices, it was known, although stated differently due to different terminology, that they minimize the Wiener index [25, 33], and maximize the number of subtrees [57].

Good trees are usually defined by a recursive or algorithmic way [39]; in this section, we present a structural characterization of them in term of semi-regularity property and show that they minimize the Γ-index among trees with a given maximum degree. The results obtained here will be used in Section 7.1 to study the minimal degree of G^n_{TBR}.

Before introducing good trees, we need some further definitions and notation. For any edge $e = (u, v)$ in T, then there is a *canonical decomposition* of T into two disjoint rooted subtrees T_u and T_v with roots

u and v, respectively. Similarly, we also associate a canonical pair of disjoint rooted trees $\{T_u, T_v\}$ with any pair $\{u, v\} \in V(T)$. On the other hand, for any vertex $v \in T$ with neighborhood $\{v_1, \ldots, v_p\}$, there exists a canonical decomposition of T (with respect to v) into p subtrees T_{v_i} rooted at v_i for each $i \in \{1, \ldots, p\}$. Unless stated otherwise, we assume in this section that all rooted subtrees of T are obtained by one of these three ways.

A tree $T \in \mathbf{R}_n^d$ is called *complete* (of height k) if $|T| = d^k$ and the d subtrees attached to the root r are all *complete* of height $k-1$. A tree $T \in \mathbf{R}_n^d$ with $n \geq d$ is called *good* of height k if among the d subtrees attached to the root, one is a good tree of height $k-1$ and the others are complete with height $k-1$ or $k-2$. Here we use the convention that a single vertex is a complete tree and a tree $T \in \mathbf{R}_n^d$ with $1 < n \leq d$ is good if and only if $h(T) = 1$. Note that if T is good (resp. complete), then every rooted subtree of T is also good (resp. complete).

Clearly, if $T \in \mathbf{R}_n^d$ is a good tree of height k, then we have $|T| \in (d^{k-1}, d^k]$. Furthermore, there is essentially (i.e., up to isomorphism) a unique good tree T in $\mathbf{R}_n^d$.

Now we have

Definition 6.2.1. *A tree $T \in \mathbf{U}_n^d$ (with $n > d$) is called* good *if there exists a vertex v in T with degree d such that all rooted subtrees in the canonical decomposition of T with respect to v are complete with height k or $k-1$ except one good rooted tree with height k.*

Note that a tree $T \in \mathbf{U}_n^d$ with $n \leq d$ is good if and only if T is a star, and that there is essentially (i.e., up to isomorphism) one good tree T in $\mathbf{U}_n^d$.

Intuitively, we get the graphs described above by simply filling the distance levels as long as there are still vertices (leaves) available, see Fig. 6.1 below.

Now we are ready to introduce the semi-regularity property. Suppose that $\{T_u, T_v\}$ is the canonical pair of rooted subtrees associated with

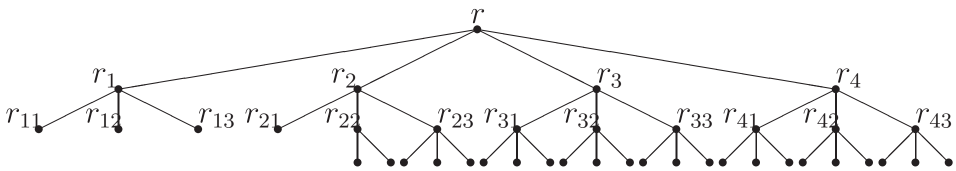

Figure 6.1: The good tree in $\mathbf{U}_{27}^{4}$.

$\{u, v\}$ in a tree $T \in \mathbf{U}_n^d$, and denote the set of subtrees attached to u in T_u by $\{T_u^1, \dots, T_u^a\}$ ($a \leq d-1$) and the set of subtrees attached to v by $\{T_v^1, \dots, T_v^b\}$ ($b \leq d-1$). Then we have the following

Definition 6.2.2. *With the above notation, $\{u, v\}$ is called* semi-regular *if one of the following two conditions holds:*

(i) $a = d-1$ *and* $\min\{|T_u^1|, \dots, |T_u^a|\} \geq \max\{|T_v^1|, \dots, |T_v^b|\}$*;*

(ii) $b = d-1$ *and* $\max\{|T_u^1|, \dots, |T_u^a|\} \leq \min\{|T_v^1|, \dots, |T_v^b|\}$.

Furthermore, a tree T in $\mathbf{U}_n^d$ is semi-regular *if each pair of its vertices is semi-regular.*

The following observations show some characteristics of a semi-regular tree. We use $\mathbf{U}_n^{d+1}$ instead of $\mathbf{U}_n^d$ in the next two results, for convenience of the notation. Note that if $T \in \mathbf{U}^{d+1}$ and T_v is a rooted subtree of T (obtained by some decompositions), then T_v belongs to $\mathbf{R}^d$.

Lemma 6.2.1. *Given a semi-regular tree T in $\mathbf{U}_n^{d+1}$, and two disjoint rooted subtrees T_u and T_v of T, then the following holds,*

(i)*: If T_u and T_v have the same height k, then $d^{k-1} < |T_u|, |T_v| \leq d^k$ holds and either T_u or T_v is complete.*

(ii)*: If $|T_u| = d^k$ for some k, then T_u is a complete tree with height k.*

(iii)*: If $d^{k-1} < |T_v| \leq |T_u| = d^k$, then T_v is a good tree with height k.*

Proof. (i): The proof is by induction; the statement is clearly true for $k = 1, 2$. For larger k, suppose that T_u and T_v both have height k. We want to show that $d^{k-1} < |T_u|, |T_v| \leq d^k$ holds and either T_u or T_v is complete. First, if neither T_u nor T_v is complete, then by induction we can assume without loss of generality that one of the following two cases occurs:

Case 1: Among the subtrees attached to the roots of T_u or T_v, there are complete subtrees with height $k-1$ and less than $k-1$;

Case 2: Among the subtrees attached to the root of T_u, there are complete subtrees with height $k-1$ and less than $k-1$, and an incomplete subtree of height $k-1$ is attached to v in T_v.

Now it is easy to verify that in both cases $\{u, v\}$ is not semi-regular, a contradiction.

By the remark of Lemma 6.2.4, $d^{k-1} < |T_u|, |T_v| \leq d^k$ follows from our induction hypothesis.

(ii): This is obvious since the complete tree is the only tree of height k in $\mathbf{R}^d$.

(iii): We establish this assertion by induction; the base cases for $k = 1, 2$ follow from the remark of Lemma 6.2.4. For larger k with $d^{k-1} < |T_v| \leq |T_u| = d^k$, we need to show that T_v is good with height k. First note that the root in T_v has d successors $\{v_1, \ldots, v_d\}$ by Lemma 6.2.4,. On the other hand, T_u is a complete tree with height k from Assertion (ii) and hence T_{u_1} is a complete tree with height $k-1$ for an arbitrary successor u_1 of u. Since $\{u, v\}$ and $\{u_1, v\}$ are both semi-regular, we have $d^{k-2} \leq |T_{v_i}| \leq d^{k-1}$ for each i. Together with Assertion (i), this implies that $h(T_{v_i}) \in \{k-2, k-1\}$ holds for each i and there are at most one incomplete subtree, say T_{v_j}, in $\{T_{v_1}, \ldots, T_{v_d}\}$ with $h(T_{v_j}) = k-1$. By induction, T_{v_j} is a good rooted subtree with height $k-1$, and hence T_v is a good tree with height k, which completes the induction. $\square$

Now we can state our main result in this section.

Theorem 6.2.2. *A tree T in $\mathbf{U}_n^{d+1}$ is semi-regular if and only if T is good.*

Proof. "$\Leftarrow$" This direction clearly holds by noting that if $\{u, v\}$ is a pair of interior vertices for a good tree T, then T_u and T_v are both good and hence they are semi-regular.

"$\Rightarrow$" Assume that $n = |T| \in (\,(d+1)d^{k-1},\ (d+1)d^k\,]$ for some $k \geq 2$ (the case for small values of k is trivial). Then there exists at least one pair of rooted subtrees of height k; by Assertion (i) in Lemma 6.2.1 one of them is complete, say T_u where $uv \in E(T)$. Now we complete the proof of this direction by considering the decomposition $T_u \cup_e T_v$ for the following two cases:

Case 1: $n \in (\,(d+1)d^{k-1},\ 2d^k\,]$. In this case, we have $|T_v| \in (d^{k-1}, d^k]$. By Assertion (iii) in Lemma 6.2.1, we know that T_v is a good tree of height k. In other words, T is good by considering the canonical decomposition of T with respect to u.

Case 2: $n \in (\,2d^k,\ (d+1)d^k\,]$. In this case we have $|T_v| \in (d^k, d^{k+1}]$ and $h(T_v) > k$. First note that by the remark of Lemma 6.2.4, we can assume that there are d successors $\{v_1, \cdots, v_d\}$ of v in T_v, and that each v_i also has d successors. Clearly we have $|T_{v_i}| \geq d^{k-1}$ and hence $h(T_{v_i}) \geq k-1$ for each i by considering the semi-regular pair $\{u, v\}$.

Now it suffices to prove the claim that $h(T_{v_i}) \leq k$ and hence $|T_{v_i}| \leq d^k$ holds for each i because together with Lemma 6.2.1, this claim implies that $h(T_{v_i}) \in \{k-1, k\}$ holds for each i and all T_{v_i} are complete except at most one, which (if exists) is good of height k.

We shall prove this claim by contradiction. Without loss of generality, assume it fails for v_1, i.e., $h(T_{v_1}) > k$; then T_{v_1} contains a subtree T_a with $h(T_a) = k$. Let b be the ancestor of a in T_{v_1}. Since $h(T_a) = h(T_u) = k$, we know $|T_a| > d^{k-1}$ from Assertion (i) in Lemma 6.2.1. By considering the semi-regular pair $\{v, b\}$, this implies $\min\{|T_{v_2}|, \ldots, |T_{v_d}|\} \geq d^k$. On the other hand, since $\{u, v_1\}$ is semi-regular, $|T_a| > d^{k-1}$ also implies that each subtree attached to v_1 in T_{v_1} has size greater than or equal to d^{k-1}, and hence $|T_{v_1}| > d^k$. Therefore we have $|T| = |T_u| + |T_{v_1}| + \cdots + |T_{v_d}| > d^{k+1}$, a contradiction as required. $\square$

We conclude this section with the following theorem, which shows that the good trees are extremal for the Γ-index.

Theorem 6.2.3. *Among trees with maximum vertex degree d and given number of leaves, precisely the good tree minimizes $\Gamma(T)$.*

Clearly, the above theorem follows directly from Theorem 6.2.2 and the following

Lemma 6.2.4. *If T is a tree in $\mathbf{U}_n^d$ with $\Gamma(T) \leq \Gamma(T')$ for any $T' \in \mathbf{U}_n^d$, then T is semi-regular.*

Proof. It suffices to show that any pair of nonleaf vertices $\{u, v\}$ in T is semi-regular. Let $u := u_0, u_1, \cdots, u_t := v$ be the unique path in T connecting u and v. To simplify notation, we put $\alpha = \alpha_1 + \cdots + \alpha_a$ with $\alpha_i := |T_u^i|$ for each i and $\beta = \beta_1 + \cdots + \beta_b$ with $\beta_j := |T_v^j|$ for each j. Without loss of generality, we can assume $\alpha_1 \leq \alpha_2 \cdots \leq \alpha_a$ and $\beta_1 \leq \beta_2 \cdots \leq \beta_b$.

For $0 < j < t$, let T_{u_j} be the subtree of T containing u_j obtained by removing the edges on $P_T(u, v)$ and put $z_j = |T_{u_j}|$,we also use the convention that $z_0 = z_t = 0$

Putting $p := p_0 + \cdots + p_t$ with $p_i := z_0 + \cdots + z_i$ and $q = q_0 + \cdots + q_t$ $q_i := z_t + \cdots + z_{t-i}$, we will prove the lemma for the case $p \leq q$, the other case is similar.

If $a + b \geq d$, let s be the $(d-1)$-th largest element in the set

$$\{\alpha_1, \cdots, \alpha_a, \beta_1, \cdots \beta_b\}$$

with respect to the order $(\mathbb{N}, \leq)$, otherwise put $s = 0$. Let $J := \{j \in [1, b] : \beta_j < s\}$ and

$$I := \{i \in [1, a] : \alpha_i \geq s \;\text{ and }\; i \geq a + b - 2 - d - |J|\}.$$

Note that $|I| + (b - |J|) \leq d - 1$ by the above definition.

For the sake of contradiction, assume that neither (i) nor (ii) in Definition 6.2.2 holds, that is, at least one of the sets I and J is not empty.

Now let T' be a tree obtained from T by reattaching every subtree T_v^j with $j \in J$ to u and reattaching every subtrees T_u^i with $i \in I$ to v. Similarly, we can define the notation α', β' for the tree T'. Note that $a+b < d$ implies $\alpha' = 0$ and $T' \in \mathbf{U}_n^d$ follows from the construction.

Since (ii) does not hold, from the construction we know $\alpha' < \min\{\alpha, \beta\}$, which implies that

$$\begin{aligned} 2\Gamma(T) - 2\Gamma(T') &= \sum_{i=0}^{t}[(\alpha + p_i)(n - \alpha - p_i) + (\beta + q_i)(n - \beta - q_i) \\ &- (\alpha' + p_i)(n - \alpha' - p_i) - (\beta' + q_i)(n - \beta' - q_i)] \\ &= \sum_{i=0}^{t}[(\alpha')^2 + (\beta')^2 - \alpha^2 - \beta^2 + \\ &\quad (\alpha - \alpha')(n - 2p_i) + (\beta - \beta')(n - 2q_i)] \\ &= 2(\alpha' - \alpha)[(\alpha' - \beta)(t + 1) + (p - q)] \\ &> 0, \end{aligned}$$

a contradiction as required. □

Remark: Following Lemma 6.2.4 is the fact that there is at most one vertex not of the maximum degree d, and such a vertex (if exists) must contain at most one nonleaf vertex in its neighborhood.

6.3 Semi-regularity principle

In this section, we will further investigate the semi-regularity property introduced in Section 6.2. We first apply it to other graphical indices, and then to the trees with a given degree sequence.

6.3.1 Applications to other graphical indices

In this subsection we will focus on trees with a given maximum degree and illustrate the idea of applying 'semi-regularity' to other graphical indices, namely the Wiener index, the number of subtrees and the weight of a tree. One can refer to Definition 6.2.2 and Lemma 6.2.4.

For the Wiener index, recall that $|V(T)|$ is the number of vertices in a tree T. Let u, v be any pair of vertices defined as in Definition 6.2.2, the same argument shows the following 'semi-regularity':

Lemma 6.3.1. *If T is a tree in $\mathbf{U}_n^d$ with $W(T) \leq W(T')$ for any $T' \in \mathbf{U}_n^d$, then one of the following must hold:*

(i) $a = d-1$ *and* $\min\{|V(T_u^1)|, \cdots, |V(T_u^a)|\} \geq \max\{|V(T_v^1)|, \cdots, |V(T_v^b)|\}$;

(ii) $b = d-1$ *and* $\max\{|V(T_u^1)|, \cdots, |V(T_u^a)|\} \leq \min\{|V(T_v^1)|, \cdots, |V(T_v^b)|\}$.

Given a tree T, a *subtree* of T is just a connected induced subgraph of T, the number of subtrees as well as related subjects were studied, see [57] and the references therein for details. Denote by $F(T)$ the number of subtrees of T and $f_T(v)$ the number of subtrees of T that contain the vertex v, then once again, we have

Lemma 6.3.2. *If T is a tree in $\mathbf{U}_n^d$ with $F(T) \geq F(T')$ for any $T' \in \mathbf{U}_n^d$, then one of the following must hold:*

(i) $a = d-1$ *and* $\min\{f_{T_u^1}(r_u^1), \cdots, f_{T_u^a}(r_u^a)\} \geq \max\{f_{T_v^1}(r_v^1), \cdots, f_{T_v^b}(r_v^b)\}$;

(ii) $b = d-1$ *and* $\max\{f_{T_u^1}(r_u^1), \cdots, f_{T_u^a}(r_u^a)\} \leq \min\{f_{T_v^1}(r_v^1), \cdots, f_{T_v^b}(r_v^b)\}$.

Here we use r_u^i (r_v^j) to denote the obvious root of the subtree T_u^i (T_v^j), the proof is a little more involved but the idea is still to reattach the branches and then compare.

Another well known index in chemistry is the *Randić index*,

$$w_\alpha = \sum_{uv \in E(T)} (\deg(u))(\deg(v))^\alpha,$$

where the sum is over all pairs of adjacent vertices and $\alpha \neq 0$. Also called the *connectivity index*, the Randić index is vigorously studied in mathematics in the recent years, see [19] and the references therein for details. When $\alpha = 1$, $w(T) := w_1(T)$ is called the *weight* of T, it is sufficient for us to work with $w(T)$ to study the extremal cases.

Using the same definition for r_u^i and r_v^j, we have the following 'semi-regularity' property:

Lemma 6.3.3. *If T is a tree in $\mathbf{U}_n^d$ with $w(T) \geq w(T')$ for any $T' \in \mathbf{U}_n^d$, then one of the following must hold:*

(i) $a = d-1$ *and* $\min\{\deg(r_u^1), \cdots, \deg(r_u^a)\} \geq \max\{\deg(r_v^1), \cdots, \deg(r_v^b)\}$*;*

(ii) $b = d-1$ *and* $\max\{\deg(r_u^1), \cdots, \deg(r_u^a)\} \leq \min\{\deg(r_v^1), \cdots, \deg(r_v^b)\}$.

This time the proof is even easier due to the nature of this concept. We only need to consider the degrees of u and v, the proof is skipped.

With Lemmas 6.3.1, 6.3.2 and 6.3.3, one can easily modify the proof of Theorem 6.2.2 to show that the corresponding extremal tree is a good tree.

6.3.2 Trees with a given degree sequence

In this subsection, we consider the extremal trees with a given degree sequence. Note that both the numbers of the vertices and the number of leaves are fixed when the degree sequence is given. The extremal trees obtained here are the same as those obtained in [60], but here we use an approach based on the following observation that analogous to the semi-regularity property.

Lemma 6.3.4. *Let T be a tree such that $\Gamma(T) \leq \Gamma(T')$ holds for all tree T' that has the same vertex degree of T. Given any path $u := u_0, u_1, \cdots, u_t := v$ with $u, v \notin L(T)$, then for the set of subtrees $\{T_u^1, \cdots, T_u^a\}$ attached to u and $\{T_v^1, \cdots, T_v^b\}$ attached to v such that $v \notin T_u^i$ and $u \notin T_v^j$ holds for each i and j, we have either*

$$a \geq b \text{ and } \min\{|T_u^1|, \cdots, |T_u^a|\} \geq \max\{|T_v^1|, \cdots, |T_v^b|\} \tag{6.2}$$

or

$$b \geq a \text{ and } \max\{|T_u^1|, \cdots, |T_u^a|\} \leq \min\{|T_v^1|, \cdots, |T_v^b|\}. \tag{6.3}$$

Proof. The proof is similar to that of Lemma 6.2.4: Let α, β, p, q be defined in the same way and we also need only to show the case $p \leq q$ here since the other one is similar. Let I and J be the two sets such that $|I| = \min\{a, b\}$, $|J| = \max\{a, b\}$,

$$I \cup J = \{T_u^1, \ldots, T_u^a, T_v^1, \ldots, T_v^b\}$$

and the number of leaves of each subtree in I is smaller than or equal to that of any subtree in J.

Now let T' be a tree obtained from T by reattaching every subtree in I to u and reattaching ever subtrees in J to v. Similarly, we can define the notation α', β' for the tree T'. Note that T' and T have the same vertex degree sequence.

Now if (6.3) does not hold, then from construction we know $\alpha' < \min\{\alpha, \beta\}$. Using an argument similar to that in Lemma 6.2.4 we have

$$2\Gamma(T) - 2\Gamma(T') = 2(\alpha' - \alpha)[(\alpha' - \beta)(t + 1) + (p - q)] > 0,$$

a contradiction as required. □

Minimization

For convenience, we will call a tree optimal if it minimizes $\Gamma(T)$ among all trees with the same degree sequence.

Consider a path in an optimal tree, after the removal of the edges on this path, some connected components will remain. Take an edge that is in the middle (or as middle as possible) of this path and label the vertices on its right as $x_1, x_2, \ldots$, and the vertices on the left as $y_1, y_2, \ldots$. Let X_i , Y_i denote the component that contains the corresponding vertex (Fig. 6.2).

We shall try to find how to arrange these components (through this operation the degree sequence stays the same) in order to minimize this index. This is the content of the next lemma.

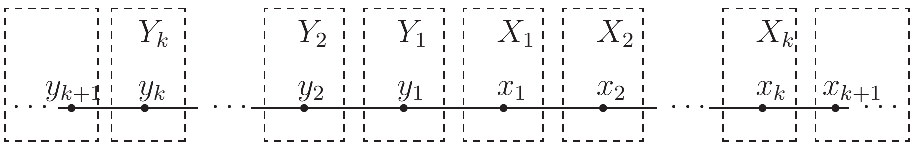

Figure 6.2: The components resulted from a path without z

Lemma 6.3.5. *In an optimal tree T, we can label the vertices such that*

$$|X_1| \geq |Y_1| \geq |X_2| \geq |Y_2| \geq \ldots \geq |X_m| = |Y_m| = 1$$

if the path is of odd length ($2m-1$); and

$$|X_1| \geq |Y_1| \geq |X_2| \geq |Y_2| \geq \ldots \geq |Y_m| = |X_{m+1}| = 1$$

if the path is of even length ($2m$).

Proof. We will prove the case when the path is of odd length, the other case is similar.

Recall that $\Gamma(T) = \sum_{e \in E} |A_e||B_e|$, since $|A_e||B_e|$ stays the same through the rearrangement for any edge e in any components X_i or Y_i, therefor

$$\begin{aligned}
\Gamma(T) &= C + \sum_{i=1}^{m} |X_i| \sum_{i=1}^{m} |Y_i| + \sum_{i=1}^{m-1} \Big((\sum_{j=1}^{i} |X_j|)(|T| - \sum_{j=1}^{i} |X_j|) \\
&\quad + (\sum_{j=1}^{i} |Y_j|)(|T| - \sum_{j=1}^{i} |Y_j|) \Big) \\
&= \sum_{1 \leq i,j \leq m} (2m+1-i-j)|X_i||Y_j| \\
&\quad + \sum_{1 \leq i < j \leq m} (j-i)(|X_i||X_j| + |Y_i||Y_j|) + C
\end{aligned} \tag{6.4}$$

where C is a constant regardless of the order of the components.

As pointed out in [60], simple application of a classic number theory result [30] yields that (6.4) is minimized when

$$|X_1| \geq |Y_1| \geq |X_2| \geq |Y_2| \geq \ldots \geq |X_m| = |Y_m| = 1.$$

$\square$

With Lemma 6.3.4, the following immediately follows, we skip the proof here:

Corollary 6.3.6. *In an optimal tree, for a path with labelling as in Lemma 6.4, we have*

$$\deg(x_1) \geq \deg(y_1) \geq \deg(x_2) \geq \deg(y_2) \geq \ldots \geq \deg(x_m) = \deg(y_m) = 1$$

if the path is of odd length ($2m-1$)*; and*

$$\deg(x_1) \geq \deg(y_1) \geq \deg(x_2) \geq \ldots \geq \deg(x_m) \geq \deg(y_m) = \deg(x_{m+1}) = 1$$

if the path is of even length ($2m$)*.*

From Corollary 6.3.6, exactly the same proof as in [60] yields:

Theorem 6.3.7. *Given the degree sequence, the greedy tree minimizes* $\Gamma(T)$*.*

While the *greedy tree* is similar to the 'good' tree, we still list its definition here for completeness, Fig. 6.3 shows a greedy tree with degree sequence $\{4, 4, 4, 3, 3, 3, 3, 3, 3, 3, 2, 2\}$.

Definition 6.3.1. *Suppose the degrees of the non-leaf vertices are given, the greedy tree is achieved by the following 'greedy algorithm':*

i) Label the vertex with the largest degree as v *(the root);*

ii) Label the neighbors of v *as* $v_1, v_2, \ldots$*, assign the largest degrees available to them such that* $\deg(v_1) \geq \deg(v_2) \geq \ldots$*;*

iii) Label the neighbors of v_1 *(except* v*) as* $v_{11}, v_{12}, \ldots$ *such that they take all the largest degrees available and that* $\deg(v_{11}) \geq \deg(v_{12}) \geq \ldots$*, then do the same for* v_2*,* v_3*,* $\ldots$*;*

iv) Repeat (iii) for all the newly labelled vertices, always start with the neighbors of the labelled vertex with largest degree whose neighbors are not labelled yet.

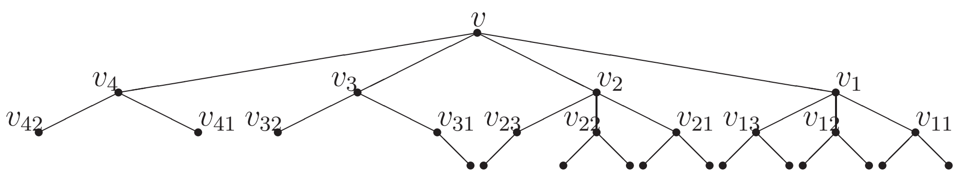

Figure 6.3: A greedy tree

Maximization

For a tree T with given degree sequence that maximizes $\Gamma(T)$, we get a similar result as Lemma 6.4 (refer to Fig. 6.2):

Lemma 6.3.8. *In a tree with a given number of vertices and degree sequence that maximizes $\Gamma(T)$, we can label the vertices on the path such that:*

$$|X_1| \leq |Y_1| \leq |X_2| \leq |Y_2| \leq \ldots \leq |X_{m-1}| \leq |Y_{m-1}|$$

if the path is of odd length $(2m-1)$; and

$$|X_1| \leq |Y_1| \leq |X_2| \leq |Y_2| \leq \ldots \leq |Y_{m-1}| \leq |X_m|$$

if the path is of even length $(2m)$.

Then again, similar arguments as in [60] lead us to:

Theorem 6.3.9. *Given the degree sequence, the greedy caterpillar maximizes $\Gamma(T)$.*

Here the *greedy caterpillar* is defined as a tree T with given degree sequence
$\{d_1 \geq d_2 \geq \ldots \geq d_k \geq 2\}$, that is formed by attaching pendant edges to a path $v_1 v_2 \ldots v_k$ of length $k-1$ such that $\deg(v_1) \geq \deg(v_k) \geq \deg(v_2) \geq$

$\deg(v_{k-1}) \geq \ldots \geq \deg(v_{[\frac{k}{2}]})$. Fig. 6.4 shows a greedy caterpillar with degree sequence $\{6, 5, 5, 5, 5, 5, 4, 3, 3\}$.

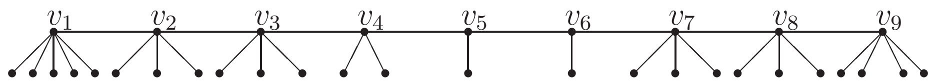

Figure 6.4: A greedy caterpillar

Chapter 7

TBR Graphs

In this chapter, we study various properties of TBR graphs, including its degree distribution and the diameter.

7.1 Degrees

In this section, we shall investigate the degrees of TBR graphs. We first establish a formula to calculate the degree of any vertex in TBR graphs by using the Γ-index introduced in Chapter 6; then apply it to obtain the maximal and minimal degrees of G^n_{TBR}. In addition, we also obtain the average degree of the nodes in the TBR graphs.

Recall that G^n_{NNI} is a regular graph with degree $2n-6$ and G^n_{SPR} is regular with degree $2(n-3)(2n-7)$. Now our main result in this section can be stated as the following

Theorem 7.1.1. *For each vertex $T \in \mathcal{T}_n$ with $n \geq 3$, we have*

$$\deg_{\text{TBR}}(T) = 4\Gamma(T) - (8n^2 - 18n + 6). \tag{7.1}$$

Let $\mathcal{O}_{\text{TBR}}(T)$ be the set of all possible TBR operations θ that can be applied to the tree T. Note that two TBR operations are distinct if either they delete different edges in the first step or the insert different edges to reconnect the two subtrees obtained from the first step. Clearly, two different TBR operations may produce the same tree, but the following lemma shows that this occurs only if both of them are NNI operations.

Lemma 7.1.2. *For any two distinct operations θ and θ' in $\mathcal{O}_{\mathrm{TBR}}(T)$, $\theta(T) = \theta'(T)$ implies that both θ and θ' are* NNI *operations.*

To prove the above lemma, we need some further definitions (see [53] for more background). Given a phylogenetic tree T with leaf set X, let $\Sigma(T)$ denote the collection of X-splits that are induced by the edges of T. Now if X' is a subset of X, then the *restriction* of T to X', denoted by $T|_{X'}$, is the unique phylogenetic tree with leaf set X' and

$$\Sigma(T|_{X'}) = \{A \cap X' | B \cap X' \ : \ A|B \in \Sigma(T) \text{ and } A \cap X' \neq \emptyset \neq B \cap X'\}.$$

Proof. Suppose that we delete some edge $e = uv$ of T in θ and $e' = u'v'$ in θ'. Then clearly we have $e \neq e'$, because otherwise $\theta(T)$ must be different from $\theta'(T)$ as the two operations are distinct.

Let $A|B$ and $A'|B'$ be the splits induced by e and e' respectively; then we may assume that $A \subset A'$ and $B' \subset B$ in view of $e \neq e'$. Suppose that in θ, the two parts $T|_A$ and T_B are reconnected by some edge f. Now, for $a \in A' - A$, we must have $T|_{A \cup a} = \theta(T)|_{A \cup a}$. Hence the edges e and f are incident with the same interior edge of $T|_A$. Following the same argument with $e' = u'v'$, we conclude that $\theta, \theta' \in \mathcal{O}_{\mathrm{SPR}}(T)$.

Without loss of generality, we may assume that $d_T(u, u') = d_T(v, v') + 2$ and let $v_0 := u, v_1 := v, v_2, \ldots, v_k := v', v_{k+1} := u'$ be the unique path of interior vertices in T joining u and u', such that $u, u' \notin \{v_1, \ldots, v_k\}$. Let $e_i = v_i v_{i+1}$, and f_i be the unique edge incident with v_i but not other v_j. Further, let C_i be the leaf set the component of $T - f_i$ that does not contain A. As $e \neq e'$, we may assume $k \geq 1$.

Note that in θ, the pruned subtree $T|_A$ cannot be grafted to any edge incident with v_1 because otherwise we have $\theta(T) = T$, contradicting the fact that $\theta \in \mathcal{O}_{\mathrm{TBR}}(T)$. We may also not attach the pruned subtree to any edge within $T|_{B'}$ in view of $A'|B' \in \Sigma(\theta(T))$.

Furthermore, Suppose that $T|_A$ is grafted to some edge contained in

$T|_{C_1}$. Then in view of $\theta(T)|_{A'} = T|_{A'}$ and $A \cup C_1 \subseteq A'$, we must have $\theta \in \mathcal{O}_{\mathrm{NNI}}(T)$. Note this establishes the lemma for the case with $k = 1$.

Now suppose $k > 1$; if we regraft $T|_A$ to any edge contained in $T|_{C_i}$ with $i > 1$, then $\theta(T)|_{A \cup C_i} \neq T|_{A \cup C_i}$, which is a contradiction. Hence we must regraft $T|_A$ to some edge in the set $\{e_2, \dots, e_k, f_2, \dots, f_k\}$.

However, the only way in which we can do this so that $\theta(T)|_{A'} = T|_A$, is if $k = 2$ and we regraft $T|_A$ to either e_2 or f_2. But then we have $\theta \in \mathcal{O}_{\mathrm{NNI}}(T)$, and hence complete the proof. $\square$

The proof of Theorem 7.1.1:

By Lemma 7.1.2, we have $\deg_{\mathrm{TBR}}(T) = |\mathcal{O}_{\mathrm{TBR}}(T)| - 3\deg_{\mathrm{NNI}}(T)$. As $\deg_{\mathrm{NNI}}(T) = 2n - 6$ holds for any $T \in \mathcal{T}_n$, it suffices to show that $|\mathcal{O}_{\mathrm{TBR}}(T)| = 4\Gamma(T) - 8(n^2 - 3n + 3)$, which is relatively straightforward.

Note that there are two types of possible TBR operations on T: the first one consists of those that induce a trivial split on T, and the second one consists of those that induce a non-trivial split. In the first case, we have n possible leaves to cut, and for each leaf x there are $2n-6$ edges in $T - x$ to which we can reconnect it so that the resulting tree is different from T.

Now let $A|B$ be some non-trivial split of T induced by the edge e. Then we we bisect T by deleting e; there are $2|A| - 3$ edge in one component of the forest and $2|B| - 3$ edges in the other. Hence, there are $(2|A| - 3)(2|B| - 3)$ ways to choose an edge from each of $T|A$ and $T|B$. Precisely one of these results in re-forming T. Hence, by taking a sum over all non-trivial splits $A|B$ of T, we get

$$\begin{aligned}
|\mathcal{O}_{\mathrm{TBR}}(T)| &= n(2n-6) + \sum_{A|B \in \Sigma_0(T)} [(2|A| - 3)(2|B| - 3) - 1] \\
&= 2n(n-3) + \sum_{A|B \in \Sigma_0(T)} [4|A||B| - 6(|A| + |B|) + 8] \\
&= 2n(n-3) + 4(\Gamma(T) - n(n-1)) - (6n-8)(n-3) \\
&= 4\Gamma(T) - 8(n^2 - 3n + 3).
\end{aligned}$$

Here in the third equality we use the fact $|\Sigma_0(T)| = n-3$ and the observation $\sum_{A|B \in \Sigma_0(T)} |A||B| = \Gamma(T) - n(n-1)$. $\square$

Theorem 7.1.3. *The tree $T \in \mathcal{T}_n$ maximizes the degree in G^n_{TBR} if and only if T is a caterpillar. In this case, we have*

$$\deg_{\mathrm{TBR}}(T) = (2n^3 - 12n^2 + 16n + 6)/3.$$

Proof. By Theorem 7.1.1, to establish the first part of the theorem it suffices to show that if $T \in \mathcal{T}_n$ is a tree such that $\Gamma(T) \geq \Gamma(T')$ holds for all $T' \in \mathcal{T}_n$, then T is a caterpillar.

Suppose that $\{x_1, x_2\}$ (resp. $\{x_3, x_4\}$) is a pair of cherries of T whose parent is u (resp. v), and $u := u_0, u_1, \cdots, u_t := v$ is the unique path $P_T(u,v)$ in T connecting u and v. For $0 < j < t$, let T_{u_j} be the subtree of T containing u_j obtained by removing the edges on $P_T(u,v)$ and put $z_j := |T_{u_j}|$. Then it suffices to show that $z_j = 1$ for $0 < j < t$.

If this fails for some $j \in \{1, \cdots, t-1\}$, we can regraft the subtree $T_{u_j} - u_j$ to the edge $x_1 u$ by a SPR operation to form a second tree T'. Now, calculating the different between $\Gamma(T)$ and $\Gamma(T')$, we find that

$$\begin{aligned}
\Gamma(T) - \Gamma(T') &= \sum_{j=0}^{i-1} (j+2)(n-j-2) - \sum_{j=0}^{i-1} (z_i + j + 1)(n - z_i - j - 1) \\
&= i(1 - z_i)(n - z_i - i - 2) \\
&< 0,
\end{aligned}$$

a contradiction as required. Note that in the last inequality we use the fact that $z_j \geq 1$ for all $j \in \{1, \cdots, t-1\}$ implies $z_i + (i-1) \leq n-4$, and hence $n - z_i - i - 2 > 0$. This is a contradiction as required.

Now it remains to calculate $\deg_{\mathrm{TBR}}(T)$ for a caterpillar T in $\mathcal{T}_n$, which is straightforward in view of Theorem 7.1.1 and the observation that $\Gamma(T) = n(n-1) + \sum_{i=2}^{n-2} i(n-i)$ holds for any caterpillar T in $\mathcal{T}_n$. $\square$

On the other hand, the following theorem presents the result for the minimal degrees of G^n_{TBR}.

Theorem 7.1.4. *The tree $T \in \mathcal{T}_n$ minimizes the degree of G^n_{TBR} if and only if T is a good tree. If the binary expansion of n is $(a_k a_{k-1} \cdots a_1 a_0)_2$, that is, we have $n = \sum_{i=0}^{k} \alpha_i 2^i$ with $\alpha_k = 1$ and $\alpha_i \in \{0, 1\}$ for $0 \leq i < k$, then for any good tree $T \in \mathcal{T}_n$, we have*

$$\begin{aligned}\deg_{\mathrm{TBR}}(T) &= 2^{k+\alpha_{k-1}}(2n - \alpha_{k-1}2^{k-1} - 2^k) - 2(4n^2 - 9n + 3) + \\ &\quad 4\sum_{j=0}^{k-2}\left(-2^j + \sum_{i=j}^{k}\alpha_i 2^i\right)\left(2n - \sum_{i=j}^{k}\alpha_i 2^i\right).\end{aligned}$$

Before presenting the proof of the above theorem, we have the following corollary, which is obtained by some straightforward calculations.

Corollary 7.1.5. *Let $T \in \mathcal{T}_n$ be a good tree; then*

$$\deg_{\mathrm{TBR}}(T) = 4n^2\lfloor \log_2 n \rfloor + O(n^2).$$

In particular, for a good tree $T \in \mathcal{T}_n$ we have

$$\deg_{\mathrm{TBR}}(T) = \begin{cases} n^2(4k - \frac{32}{3}) + 22n - 6 & \text{if } n = 3 \cdot 2^{k-1} \text{ for some } k, \\ n^2(4k - 13) + 22n - 6 & \text{if } n = 2^k \text{ for some } k. \end{cases}$$

□

The proof of Theorem 7.1.4

The first part of the theorem clearly follows from Theorem 6.2.3 and Theorem 7.1.1. To establish the second part, we put $\beta_j := \frac{1}{2^j}\sum_{i=j}^{k}\alpha_i 2^i$, and note from the definition that there are exactly β_j distinct subtrees of height j in T, among which all are of size 2^j (i.e., they are complete) with at most one exception, which has size $n - 2^j(\beta_j - 1)$.

Now we shall prove the theorem by considering the following two cases:

Case 1: $\alpha_{k-1} = 1$. In this case, there exists a canonical one-one and onto correspondence between the subtrees of height t with $0 \leq t \leq k-1$

and edges in T. Therefore, we have

$$\begin{aligned}\Gamma(T) &= \sum_{j=0}^{k-1}[(\beta_j-1)(2^j-1)(n-2^j)+(n-2^j(\beta_j-1))2^j(\beta_j-1)\\ &= \sum_{j=0}^{k-1}2^j(\beta_j-1)(2n-2^j\beta_j)\\ &= \sum_{j=0}^{k-1}\left(-2^j+\sum_{i=j}^{k}\alpha_i 2^i\right)\left(2n-\sum_{i=j}^{k}\alpha_i 2^i\right).\end{aligned}$$

Together with (7.1), this completes the proof of this case.

Case 2: $\alpha_{k-1}=0$. The proof of this case is similar to the first one. But we need to note that in this case we have $\beta_{k-1}=2$ and hence the two subtrees of height $k-1$ are mapped to the same interior edge in T in the canonical correspondence. □

We conclude this section with a brief discussion on the average degree of the nodes in the TBR graph. In other words, if a tree $T \in \mathcal{T}_n$ is generated by the *uniform model*, where each tree in $\mathcal{T}_n$ is chosen with equal probability, then it is not difficult to obtain the following theorem by previous results on the expected distance between leaves.

Theorem 7.1.6. *Let T_n be a random tree in $\mathcal{T}_n$ generated by the uniform model; then we have $\mathbb{E}(\deg_{\mathrm{TBR}}(T_n)) \sim 2\sqrt{\pi}n^{5/2}$.*

Proof. Following Theorem 3.1 in [56], the expected distance between a pair of leaves in the tree T_n is asymptotic to $\sqrt{\pi n}$, and hence we have $\mathbb{E}(\Gamma(T_n)) \sim \sqrt{\pi}n^{5/2}/2$, which implies the theorem in view of (7.1). □

7.2 The Yule-Harding model

In this section, we study the distribution of the size of the unit-neighborhood of trees generated by the Yule-Harding model.

Denoting the set of rooted binary phylogenetic trees with n leaves by $\mathcal{T}_n^*$, this model generates a random element of $\mathcal{T}_n^*$ as follows. Starting with a subtree with just three leaves that are randomly labelled by three distinct elements in $\{1, \ldots, n\}$, recursively select a random pendant edge with uniform probability and make the next leaf, which is labelled by choosing with uniform probability one of the labels from $\{1, \ldots, n\}$ that does not used so far, adjacent to the midpoint of that edge. This procedure stops when the resulting tree has n leaves.

This model has been widely studied and has many attractive properties. Note that if we suppress the root in the trees generated by the Yule-Harding model, then we can also regard it generates a random element in $\mathcal{T}_n$.

Now we can state our main result in this section.

Theorem 7.2.1. *Let T_n be a random element in $\mathcal{T}_n$ generated by the Yule-Harding model and let D_n be the random variable defined as $\deg_{\mathrm{TBR}}(T_n)$; then we have*

$$\mathbb{E}(D_n) = 8n(n+1)H_n - \frac{70n^2 - 56n + 18}{3} \sim 8n^2 \ln(n)$$

and

$$\mathrm{Var}(D_n) \sim \frac{11284 - 480\pi^2}{45} n^4,$$

where $H_n := \sum_{j=1}^n 1/j$ denotes the n-th harmonic number.

In order to establish the above theorem, we shall first consider the Γ-index for rooted trees, which is related to that of unrooted trees by the following observation: For any rooted tree $T^* \in \mathcal{T}^*$, let T_1^* and T_2^* be the left and right subtree of the root of T^*; then we have

$$\Gamma(T^*) = \Gamma(T) + |T_1^*| \cdot |T_2^*|, \tag{7.2}$$

where T is the tree obtained from T^* by suppressing the root r.

Furthermore, the *Sackin index* of a rooted tree T^* is defined as

$$S(T^*) := \sum_{u \in L(T^*)} \mathrm{dist}_{T^*}(u, r).$$

Now we have

$$S(T^*) = S(T_1^*) + S(T_2^*) + n \tag{7.3}$$

and

$$\Gamma(T^*) = \Gamma(T_1^*) + \Gamma(T_2^*) + b(T_1^*, T_2^*), \tag{7.4}$$

where $b(T_1^*, T_2^*) := 2|T_1^*||T_2^*| + |T_1^*|S(T_2^*) + |T_2^*|S(T_1^*)$. It is known (cf. [38]) that for a random tree T_n^* in $\mathcal{T}_n^*$ generated by the Yule-Harding model, we have

$$\mathbb{E}(S(T_n^*)) = 2n(H_n - 1), \tag{7.5}$$

where $H_n := \sum_{j=1}^{n} 1/j = \ln(n) + O(1)$ denotes the n-th harmonic number.

To investigate the Γ-index of a random Yule-Harding tree, we need to introduce some further definitions and notation. We denote by A^t the transpose of a vector or matrix A; by $\stackrel{\mathcal{D}}{=}$ the equality in distribution of the left and right hand side; by $\mathcal{L}(\mathcal{X})$ the distribution of X; by $X_n \xrightarrow{\mathcal{D}} X$ the convergence of $\mathcal{L}(X_n)$ to $\mathcal{L}(X)$. Finally, let $\mathcal{M}_2$ be the space of all centered probability measures on $\mathbb{R}^2$ with finite second moments.

Now we can state the following theorem, a key step to establish Theorem 7.2.1.

Theorem 7.2.2. *Let* (Γ_n^*, S_n^*) *denote the vector of the* Γ *and Sackin index of a random tree in* $\mathcal{T}_n^*$ *generated by the Yule-Harding model; then* *we have*

$$
\begin{aligned}
\mathbb{E}(\Gamma_n^*) &= 2n(n+1)H_n - 4n^2,\\
\operatorname{Var}(\Gamma_n^*) &\sim \frac{188-6\pi^2}{9}n^4,\\
\operatorname{Cov}(\Gamma_n^*, S_n^*) &\sim \frac{68-6\pi^2}{9}n^3,\\
\operatorname{Cor}(\Gamma_n^*, S_n^*) &\sim \frac{68-6\pi^2}{\sqrt{63-6\pi^2}\sqrt{188-6\pi^2}}\\
\left(\frac{\Gamma_n^* - \mathbb{E}\Gamma_n^*}{n^2}, \frac{S_n^* - \mathbb{E}S_n^*}{n}\right) &\xrightarrow{\mathcal{D}} (\Gamma^*, S^*),
\end{aligned}
$$

where $\mathcal{L}(\Gamma^*, S^*)$ is the unique fixed-point of the map $\mathbb{T}: \mathcal{M}_2 \to \mathcal{M}_2$ given for $\nu \in \mathcal{M}_2$ by

$$
\mathbb{T}(\nu) := \mathcal{L}\left(\begin{bmatrix} U^2 & U(1-U) \\ 0 & U \end{bmatrix}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} + \begin{bmatrix} (1-U)^2 & U(1-U) \\ 0 & 1-U \end{bmatrix}\begin{pmatrix} Z_1' \\ Z_2' \end{pmatrix} + \begin{pmatrix} \widehat{b}_1 \\ \widehat{b}_2 \end{pmatrix}\right),
$$

with

$$
\begin{pmatrix} \widehat{b}_1 \\ \widehat{b}_2 \end{pmatrix} := \begin{pmatrix} 2U\ln U + 2(1-U)\ln(1-U) - 10U(1-U) \\ 2U\ln U + 2(1-U)\ln(1-U) + 1 \end{pmatrix},
$$

where (Z_1, Z_2), (Z_1', Z_2'), U are independent with $\mathcal{L}(Z_1, Z_2) = \mathcal{L}(Z_1', Z_2') = \nu$ and U uniform $[0,1]$ distributed.

The main technique used in the following proof is a multivariate contraction method developed by Neininger [46], which have been used to study other graphical indices on random trees [9, 47].

Proof. Let T_n^* be a rooted Yule-Harding random tree with n leaves; then $\Gamma_n^* := \Gamma(T_n^*)$ and $S_n^* := S(T_n^*)$. Denote by T_1^* and T_2^* the left and right subtree of T_n^* attached to the root, and put $I_n := |T_1^*|$ and $J_n := |T_2^*|$. We begin by noting that $I_n + J_n = n$ and I_n is a uniform distributed random variable over $\{1, 2, \dots, n-1\}$ (cf [54]).

Next, by (7.3) and (7.4), we have the following two recurrences on distributions:

$$S_n^* \stackrel{\mathcal{D}}{=} S_{I_n}^* + (S_{J_n}^*)' + n \tag{7.6}$$

and

$$\Gamma_n^* \stackrel{\mathcal{D}}{=} \Gamma_{I_n}^* + (\Gamma_{J_n}^*)' + b_n \tag{7.7}$$

with (Γ_n, S_n), $((\Gamma_n^*)', (S_n^*)')$ and I_n being independent and

$$b_n := 2|I_n||J_n| + |J_n|S_{I_n}^* + |I_n|(S_{J_n}^*)'.$$

Here the initial conditions for the above recurrences are $\Gamma_1^* = b_1 = 0$ and $\Gamma_2^* = b_2 = 2$.

Now we shall divide the remainder of the proof into three steps.

Step 1: Expectation. To simplify the notation, put $\alpha_n := \mathbb{E}(\Gamma_n^*)$ and $\beta_n := \mathbb{E}(b_n)$; then from (7.7) and the distribution of I_n, we have

$$\alpha_n = \beta_n + \frac{2}{n-1}\sum_{k=1}^{n-1} \alpha_k \qquad (n \geq 2)$$

with $\alpha_1 = \beta_1 = 0$ and $\alpha_2 = \beta_2 = 2$. Solving the above recurrence (see for example [27]), we can conclude that

$$\alpha_n = \beta_n + 2n\sum_{k=2}^{n-1} \frac{\beta_k}{k(k+1)}.$$

holds for $n \geq 2$, which yields the first assertion of the theorem since

$$
\begin{aligned}
\beta_m &= \mathbb{E}(b_m) \\
&= \frac{1}{m-1}\sum_{k=1}^{m-1} 2k(m-k) + (m-k)\mathbb{E}(S_k) + k\mathbb{E}(S_{m-k}) \\
&= \frac{2}{m-1}\sum_{k=1}^{m-1} k(m-k)(H_k + H_{m-k} - 1)) \\
&= \frac{4m}{m-1}\sum_{k=1}^{m-1} kH_k - \frac{4}{m-1}\sum_{k=1}^{m-1} k^2 H_k - \frac{2}{m-1}\sum_{k=1}^{m-1} k(m-k) \\
&= \frac{2m(m+1)}{3}H_m - \frac{m(8m+2)}{9}
\end{aligned}
$$

holds for $m \geq 2$. Here we refer the reader to [55, Table 1] for the formulas used in the last equality to calculate $\sum_{k=1}^{m} kH_k$ and $\sum_{k=1}^{m} k^2 H_k$.

Step 2: Limit laws. Now considering a rescaled version of the Sackin and Γ-index:

$$
\overline{S}^*_n := \frac{S^*_n - E(S^*_n)}{n} \quad \text{and} \quad \overline{\Gamma}^*_n := \frac{\Gamma^*_n - E(\Gamma^*_n)}{n^2},
$$

and putting

$$
X_n := \begin{pmatrix} \overline{S}^*_n \\ \overline{\Gamma}^*_n \end{pmatrix} \text{ and } X'_n := \begin{pmatrix} (\overline{S}^*_n)' \\ (\overline{\Gamma}^*_n)' \end{pmatrix},
$$

then from (7.6), (7.7) and (7.5), we have

$$
X_n \overset{\mathcal{D}}{=} A^n_1 X_{I_n} + A^n_2 X'_{J_n} + c^n,
$$

where

$$
A^n_1 := \begin{pmatrix} I_n^2/n & I_n J_n/n^2 \\ 0 & I_n/n \end{pmatrix}, \quad A^n_2 := \begin{pmatrix} J_n^2/n & I_n J_n/n^2 \\ 0 & J_n/n \end{pmatrix}, \quad c^n := \begin{pmatrix} c^n_1 \\ c^n_2 \end{pmatrix}
$$

with

$$
\begin{aligned}
c_1^n &= \frac{I_n\mathbb{E}((S^*_{J_n})') + J_n\mathbb{E}(S^*_{I_n}) + 2I_nJ_n + \alpha_{I_n} + \alpha_{J_n} - \alpha_n}{n^2} \\
&= \frac{2I_n}{n}\ln\frac{I_n}{n} + \frac{2J_n}{n}\ln\frac{J_n}{n} - \frac{10I_nJ_n}{n^2} + o(1)
\end{aligned}
$$

and

$$
c_2^n = \frac{n + \mathbb{E}(S^*_{I_n}) + \mathbb{E}((S^*_{J_n})') - \mathbb{E}(S^*_n)}{n} = \frac{2I_n}{n}\ln\frac{I_n}{n} + \frac{2J_n}{n}\ln\frac{J_n}{n} + 1 + o(1).
$$

Therefore, by dominated converge, we obtain the following convergences in L_2:

$$
A_1^n \;\to\; \widehat{A}_1 := \begin{bmatrix} U^2 & U(1-U) \\ 0 & U \end{bmatrix}, \tag{7.8}
$$

$$
A_2^n \;\to\; \widehat{A}_2 := \begin{bmatrix} (1-U)^2 & U(1-U) \\ 0 & (1-U) \end{bmatrix}, \tag{7.9}
$$

$$
b^n \;\to\; \widehat{b} := \begin{pmatrix} 2U\ln U + 2(1-U)\ln(1-U) - 10U(1-U) \\ 2U\ln U + 2(1-U)\ln(1-U) + 1 \end{pmatrix} \tag{7.10}
$$

where U denotes a random variable distributed uniform on $[0,1]$.

Then the multivariate contraction theorem [46] claims that the sequence (X_n) converges in distribution and with second moments to a distribution $\mathcal{L}(X)$, which is the unique fixed-point of the map $\mathbb{T} : \mathcal{M}_2 \to \mathcal{M}_2$ given by

$$
\mathbb{T}(\nu) := \mathcal{L}(\widehat{A}_1 Z + \widehat{A}_2 Z' + \widehat{b}), \tag{7.11}
$$

with $(\widehat{A}_1, \widehat{A}_2, \widehat{b}), Z, Z'$ are independent and $\mathcal{L}(Z) = \mathcal{L}(Z') = \nu$, if the following conditions are satisfied:

(i) $(A_1^{(n)}, A_2^{(n)}, b^{(n)}) \xrightarrow{L_2} (\widehat{A}_1, \widehat{A}_2, \widehat{b}), \quad n \to \infty,$

(ii) $\mathbb{E}\left[||(\widehat{A}_1)^t \widehat{A}_1||_{op}\right] + \mathbb{E}\left[||(\widehat{A}_2)^t \widehat{A}_2||_{op}\right] < 1,$

(iii) $\mathbb{E}\left[\mathbf{1}_{\{I_n \le l\}} ||(A_1^{(n)})^t A_1^{(n)}||_{op}\right] \to 0,$ for all $l \in \mathbb{N},\ n \to \infty,$

(iv) $\mathbb{E}\left[\mathbf{1}_{\{I_n\le l\}}||(A_2^{(n)})^t A_2^{(n)}||_{op}\right] \to 0,$ for all $l\in\mathbb{N},\ n\to\infty,$

where $||A||_{op} = \sup_{||x||=1}||Ax||$ denotes the operator norm of A and $\mathbf{1}_B$ denotes the indicator function of a set B.

Therefore, to complete this step, it suffices to verify the above conditions: Indeed, (i) follows directly from (7.8), (7.9) and (7.10); (iii) and (iv) hold since $||(A_r^{(n)})^t A_r^{(n)})||_{op}$ are deterministically bounded ($r = 1, 2$) and

$$\mathbb{P}(\{I_n \le l\}) = \mathbb{P}(\{J_n \le l\}) \le \frac{l}{n} \to 0$$

holds for all $l \in \mathbb{N}$ and $n \to \infty$. Finally, we shall check (ii): clearly the largest eigenvalue $\lambda(U)$ in absolute value for $(\widehat{A}_1)^t\widehat{A}_1$ is

$$\lambda(U) = U^2\left(\frac{1+U^2+(1-U)^2}{2} + \sqrt{\frac{(1+U^2+(1-U)^2)^2}{4} - U^2}\right),$$

which implies that

$$\begin{aligned}
\mathbb{E}\left[||(\widehat{A}_1)^t\widehat{A}_1||_{op}\right] + \mathbb{E}\left[||(\widehat{A}_1)^t\widehat{A}_1||_{op}\right] &= 2\mathbb{E}\left[\lambda(U)\right] \\
&= \frac{3}{10} + \frac{29}{60}\sqrt{2} + \frac{1}{4}\ln(\sqrt{2}-1) \\
&< 1
\end{aligned}$$

since $(\widehat{A}_1)^t\widehat{A}_1$ and $(\widehat{A}_2)^t\widehat{A}_2$ are identically distributed.

Step 3: Second Moments. To simplify the notation, put $\varepsilon(U) := U\ln U + (1-U)\ln(1-U)$. From Step 2, equation (7.11) has a unique solution, so we can choose two independent copies (Γ^*, S^*) and $((\Gamma^*)', (S^*)')$ with $\mathcal{L}(\Gamma^*, S^*) = \mathcal{L}((\Gamma^*)', (S^*)')$ being the fixed-point of $\mathbb{T}$ in $\mathcal{M}_2$. Then $\mathbb{E}\Gamma^* = \mathbb{E}S^* = 0$ and

$$\begin{pmatrix} \Gamma^* \\ S^* \end{pmatrix} \overset{\mathcal{D}}{=} \begin{pmatrix} U^2\Gamma^* + U(1-U)(S^*+(S^*)') + (1-U)^2(\Gamma^*)' - 10U(1-U) + 2\varepsilon(U) \\ US^* + (1-U)(S^*)' + 1 + 2\varepsilon(U) \end{pmatrix}.$$

Furthermore, together with the independence property and $\mathbb{E}\Gamma^* =$

$\mathbb{E}S^* = 0$, the above equality implies

$$\begin{aligned}\mathbb{E}[\Gamma^* S^*] &= \mathbb{E}[U^3\Gamma^* S^*] + \mathbb{E}[(1-U)^3(\Gamma^*)'(S^*)'] + \mathbb{E}[U^2(1-U)(S^*)^2] \\ &\quad + \mathbb{E}[U(1-U)^2((S^*)')^2] + \mathbb{E}[(2\varepsilon(U) - 10U(1-U))(1+2\varepsilon(U))]\end{aligned}$$

Using the fact that $\mathbb{E}((S^*)^2) = 7 - 2\pi^2/3$ (cf. [51, 8]), we obtain

$$\mathrm{Cov}(\Gamma^*, S^*) = \mathbb{E}[\Gamma^* S^*] = \frac{68 - 6\pi^2}{9},$$

which will be use to calculate $\mathbb{E}[(\Gamma^*)^2]$. We have

$$\begin{aligned}\mathbb{E}[(\Gamma^*)^2] &= \mathbb{E}[(1-U)^4((\Gamma^*)')^2] + 2\mathbb{E}[U^3(1-U)\Gamma^* S^*] + \mathbb{E}[U^4(\Gamma^*)^2] \\ &\quad + 2\mathbb{E}[U(1-U)^3(\Gamma^*)'(S^*)'] + \mathbb{E}[U^2(1-U)^2(S^* + (S^*)')^2] \\ &\quad + \mathbb{E}[(-10U(1-U) + 2\varepsilon(U))^2],\end{aligned}$$

which implies

$$\mathbb{E}[(\Gamma^*)^2] = \frac{188 - 6\pi^2}{9}, \quad \text{and} \quad \mathrm{Cor}(\Gamma^*, S^*) = \frac{68 - 6\pi^2}{\sqrt{63 - 6\pi^2}\sqrt{188 - 6\pi^2}}$$

Since the convergence

$$\left(\frac{\Gamma_n^* - \mathbb{E}(\Gamma_n^*)}{n^2}, \frac{S_n^* - \mathbb{E}(S_n^*)}{n}\right) \xrightarrow{\mathcal{D}} (\Gamma^*, S^*)$$

holds with second moments, this implies

$$\mathrm{Var}(\Gamma_n^*) \sim n^4 \mathrm{Var}(\Gamma^*), \ \mathrm{Cov}(\Gamma_n^*, S_n^*) \sim n^3 \mathrm{Cov}(\Gamma^*, S^*)$$

and $\mathrm{Cor}(\Gamma_n^*, S_n^*) \sim \mathrm{Cor}(\Gamma^*, S^*)$. □

With the above theorem, we are ready to prove Theorem 7.2.1.

Proof of Theorem 7.2.1:

We shall use the same notation T_n^*, T_1^*, T_2^*, I_n, Γ_n^* and S_n^* as defined in the proof of Theorem 7.2.2. In addition, let T_n be the unrooted tree obtained from T_n^* by suppressing the root; put $\Gamma_n := \Gamma(T_n)$ and $Y_n := |T_1^*| \cdot |T_2^*|$.

First, from the distribution of I_n, we can assert that

$$\mathbb{E}(Y_n) = \frac{1}{n-1}\sum_{i=1}^{n-1} i(n-i) = \frac{n^2+n}{6} \tag{7.12}$$

and

$$\mathrm{Var}(Y_n) = \mathbb{E}(Y_n^2) - (\mathbb{E}(Y_n))^2 = \frac{1}{30}n^4 + O(n^3). \tag{7.13}$$

Next, we also have

$$\begin{aligned}
\mathbb{E}(\Gamma_n^* Y_n) &= \frac{1}{n-1}\sum_{i=1}^{n-1}\mathbb{E}(\Gamma_n^* Y_n | I_n = i) \\
&= \frac{4(n+1)}{n-1}\sum_{i=1}^{n-1} i^2(n-i)H_i - \frac{2}{n-1}\sum_{i=1}^{n-1}[i^2(n-i)^2 - 4i^3(n-i)] \\
&= \frac{n^4 H_n}{3} - \frac{119}{180}n^4 + O(n^3),
\end{aligned}$$

where in the last equality we use the fact that $\sum_{i=1}^{n-1} i^2 H_i = n^3 H_n/3 - n^3/9 + o(n^3)$ and $\sum_{i=1}^{n-1} i^3 H_i = n^4 H_n/4 - n^4/16 + o(n^4)$ while the second equality holds because the recurrence relation (7.4) implies

$$\begin{aligned}
\mathbb{E}(\Gamma_n^* Y_n | I_n = i) &= i(n-i)\mathbb{E}(\Gamma_n^* | I_n = i) \\
&= i(n-i)\mathbb{E}(\Gamma_i^* + \Gamma_{n-i}^* + 2i(n-i) + iS_{n-i}^* + (n-i)S_i^*) \\
&= 2i^2(n-i)^2(H_i + H_{n-i} - 1) - 4i(n-i)(i^2 + (n-i)^2) + \\
&\qquad 2i^2(i+1)(n-i)H_i + 2i(n-i)^2(n-i+1)H_{n-i}
\end{aligned}$$

for $i \in \{1, \ldots, n\}$. Here we use (7.5) and Theorem 7.2.2 to obtain the last equality.

Finally, since $\Gamma_n = \Gamma_n^* - Y_n$ holds in view of (7.2), from Theorem 7.2.2 we have

$$\mathbb{E}(\Gamma_n) = \mathbb{E}(\Gamma_n^*) - \mathbb{E}(Y_n) = 2n(n+1)H_n - \frac{23n^2 - n}{6}$$

and

$$\begin{aligned}
\mathrm{Var}(\Gamma_n) &= \mathrm{Var}(\Gamma_n^*) + \mathrm{Var}(Y_n) + \mathbb{E}(\Gamma_n^* Y_n) - \mathbb{E}(\Gamma_n^*)\mathbb{E}(Y_n) \\
&\sim \frac{188 - 6\pi^2}{9} n^4 + \frac{61}{180} n^4 \\
&= \frac{2821 - 120\pi^2}{180} n^4.
\end{aligned}$$

Together with (7.1), this yields the theorem. □

7.3 Diameter

Given a graph G, its *diameter* $\mathrm{diam}(G)$ is defined to be the maximal distance between all pairs of vertices in G. In this section, we obtain a current best known lower bound on the diameter of G^n_{TBR}. Note that the best known upper bound is $n - \lfloor \sqrt{n}/2 \rfloor$ and a lower bound similar to the one presented here, $n - 2\lceil \sqrt{n}\, \rceil + 1$, is independently obtained by Stefan Grunewald [28] with a more complicated approach using the technique of agreement forest.

Theorem 7.3.1. *Suppose that $n \in [k^2, (k+1)^2)$ holds for for some positive number $k \geq 2$; then we have*

$$\mathrm{diam}(G^n_{\mathrm{SPR}}) \geq \mathrm{diam}(G^n_{\mathrm{TBR}}) \geq \begin{cases} n - 2\sqrt{n} + 1 & \text{if } n = k^2; \\ n - 2k & \text{if } n > k^2. \end{cases}$$

To establish the above theorem, we need some additional concepts from phylogenetics. Recall that a *character* f on X is a surjective map from X to a finite set $\mathcal{C} = \{1, \cdots, r\}$, which is called the *state set* of f and whose size $|\mathcal{C}|$ will be denoted by $|f|$. An *extension* of f to a tree T with $L(T) = X$ is a function $\bar{f} : V(T) \to \mathcal{C}$ such that $f(x) = \bar{f}(x)$ holds for any leaf $x \in L(T)$; the *changing number* of $\bar{f}$, denoted by $ch(\bar{f})$, is the number of edges $\{u, v\}$ with $\bar{f}(u) \neq \bar{f}(v)$. Now given a pair (T, f) such that f is a character on $L(T)$, the *parsimony score* of f on T, denoted by $l(T, f)$, is defined as

$$\min\{ch(\bar{f}) \,:\, \bar{f} \text{ is an extension of } f \text{ to } T\},$$

while the *homoplasy score* of (T, f) is defined as

$$h(T, f) := l(T, f) - |f| + 1.$$

Now we can state the following lemma, a key step to prove the above theorem.

Lemma 7.3.2. *For any $n \geq 4$, we have*

$$\max_{T \in \mathcal{T}_n} \max_{f} h(T, f) \leq \operatorname{diam}(G^n_{\mathrm{TBR}}) \leq \operatorname{diam}(G^n_{\mathrm{SPR}}).$$

Proof. The second inequality is trivial. To establish the first one, we use a result by Bryant [12]: if T differs from T' by a single TBR operation, then $l(f, T') \leq l(f, T) + 1$ holds for any character f on $L(T)$. This implies that the distance between T and T' in G^n_{TBR} is bounded below by $h(T, f)$ for any character f with $h(T', f) = 0$. Now consider a pair (T, f) such that $h(T, f)$ obtained the maximal value in the left side; then the inequality clearly holds because there always exists some tree T' in $\mathcal{T}_n$ with $h(T', f) = 0$. $\square$

With the above lemma, we can obtain a lower bound on $\operatorname{diam}(G^n_{\mathrm{TBR}})$ by choosing a suitable pair (T, f), as indicated by the following proof.

The proof of Theorem 7.3.1:

If $n = k^2$, it suffices to construct a tree T and a character f such that $h(T, f) = (k-1)^2$ in view of Lemma 7.3.2. Consider a caterpillar tree on $\{1, 2, \cdots, n\}$ such that $\{1, 2\}$ and $\{n-1, n\}$ are the only two cherries and the other leaves are labelled consecutively from 3 to $n-2$.

Let f be the character on X with defined as

$$f(x) := \quad x \mod k \qquad \text{for } x \in \{1, 2, \cdots n\}.$$

Note that the state set of f is $\mathbb{Z}_k$ and hence $|f| = k$; then it is clear that we have $l(T, f) = (k-1)k$, and therefore $h(T, f) = (k-1)^2$.

If $k^2 < n < (k+1)^2$, then we can consider the pair (T, f) constructed as above and a similar analysis shows $l(T, f) = k(k-1) + (n - k^2 - 1)$, and hence $h(T, f) = n - 2k$. $\square$

Note that the bound obtained in Theorem 7.3.1 is "optimal" with respect to the approach in this section. More precisely, we have

$$\max_{T \in \mathcal{T}_n} \max_f h(T, f) \leq n - 2\sqrt{n} + 1$$

for any given n: Indeed, consider a pair (T, f) such that $h(T, f)$ is maximal and put $r := |f|$; then there exists a state $\alpha \in f(X)$ such that $|f^{-1}(\alpha)| \geq n/r$ holds, which implies $l(T, f) \leq n(1 - 1/r)$, and hence

$$h(T, f) \leq n(1 - \frac{1}{r}) - (r - 1) = (n + 1) - (r + \frac{n}{r}) \leq (n + 1) - 2\sqrt{n}.$$

Bibliography

[1] D. Aldous. Probability distributions on cladogram. In D. Aldous and R. Pemantl, editors, *Random Discrete Structures*, pages 1–18. Springer, Berlin, 1996.

[2] D. Aldous. Stochastic models and descriptive statistics for phylogenetics trees, from yule to today. *Statistical Science*, 16:23–34, 2001.

[3] B. Allen and M. Steel. Subtree transfer operations and their induced metrics on evolutionary trees. *Annals of Combinatorics*, 5:1–15, 2001.

[4] J. Bang-Jensen and G. Gutin. *Digraphs. Theory, Algorithms and Applications*. Springer, 2000.

[5] J.-C. Bermond, F. Comellas, and D. F. Hsu. Distributed loop computer networks: A survey. *Journal of Parallel and Distributed Computing*, 24(1):2–10, 1995.

[6] J. Berstel and J. Karhumäki. Combinatorics on words - a tutorial. *Bulletin of the EATCS*, 79:178–228, 2003.

[7] N. Biggs. *Interaction Models*, volume 30 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 1977.

[8] M. Blum and O. Francois. On statisitcal tests of phylogenetic tree imbalance: The sackin and other indices revisited. *Mathematical Biosciences*, 195:141–153, 2005.

[9] M. Blum, O. Francois, and S. Janson. The mean, variance and limiting distribution of two statistics sensitive to phylogenetic tree balance. *Annals of Applied Probability*, 16(4):2195–2214, 2006.

[10] F. Boesch and R. Tindell. Circulants and their connectivities. *Journal of Graph Theory*, 8:487–499, 1984.

[11] J. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30, 1965.

[12] D. Bryant. The splits in the neighborhood of a tree. *Annals of Combinatorics*, 8:1–11, 2004.

[13] J. Cai, G. Havas, B. Mans, A. Nerurkar, J. Seifert, and I. Shparlinski. On routing in circulant graphs. In *COCOON: Annual International Conference on Computing and Combinatorics*, 1999.

[14] P. J. Cameron and T. Wu. A new characterization of balanced words. In N. Bédaride P. Arnoux and J. Cassaigne, editors, *WORDS2007: Proceedings of the 6th International Conference on Words*, pages 63–71, 2007.

[15] P. J. Cameron and T. Wu. Regular configurations via the Kawasaki-Ising model. *Preprint*, 2007.

[16] J. Clough and J. Douthett. Maximally even sets. *Journal of Music Theory*, 35(1–2):93–173, 1991.

[17] J. Clough and G. Myerson. Music scales and the generalized cycle of fifths. *The American Mathematical Monthly*, 93(9):695–701, 1986.

[18] B. Codenotti, I. Gerace, and S. Vigna. Hardness results and spectral techniques for combinatorial problems on circulant graphs. *Linear Algebra and its Applications*, 285(1–3):123–142, 1998.

[19] C. Delorme, O. Favaron, and D. Rautenbach. On the randić index. *Discrete Mathematics*, 257:29–38, 2002.

[20] R. Diestel. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, 1997.

[21] A. Dobrynin, R. Entringer, and I. Gutman. Wiener index of trees: theory and applications. *Acta Applicandae Mathematicae*, 66(3):211–249, 2001.

[22] J. Douthett and R. Krantz. Energy extremes and spin configurations for the one-dimensional anti-ferromagnetic ising model with arbitrary-range interaction. *Journal of Mathematical Physics*, 37(7):3334–3353, 1996.

[23] J. Douthett and R. Krantz. Maximally even sets and configurations: common threads in mathematics, physics, and music. *Journal of Combinatorial Optimization*, 223(1–3):385–410, 2007.

[24] B. Elspas and J. Turner. Graphs with circulant adjacency matrices. *Journal of Combinatorial Theory*, 9:297–307, 1970.

[25] M. Fischermann, A. Hoffmann, D. Rautenbach, L. Székely, and L. Volkmann. Wiener index versus maximum degree in trees. *Discrete Applied Mathematics*, 122(1–3):127–137, 2002.

[26] F. Göbel and E. A. Neutel. Cyclic graphs. *Discrete Applied Mathematics*, 99(1-3):3–12, 2000.

[27] R. Graham, D. Knuth, O., and Patashnik. *Concrete Mathematics*. Addison Wesley, 1993.

[28] S. Grunewald. Personal communication. 2008.

[29] E. Harding. The probabilities of rooted tree-shapes generated by random bifurcation. *Advances in Applied Probability*, 3:44–77, 1971.

[30] G. Hardy. *Inequalities*. Cambridge, 1934.

[31] P. Humphries. Bounds on the size of the tbr unit-neighborhood. *Annals of Combinatorics*, to appear.

[32] P. Humphries and T. Wu. The unit-neighborhood of tree rearrangement operations. *preprint*, 2008.

[33] F. Jelen and E. Triesch. Superdominance order and distance of trees with bounded maximum degree. *Discrete Applied Mathematics*, 125(2–3):225–233, 2003.

[34] O. Jenkinson. Balanced words and majorization. *Preprint*, 2006.

[35] O. Jenkinson. Optimization and majorization of invariant measures. *Electronic Research Announcements of the American Mathematical Society*, 13:1–12, 2007.

[36] O. Jenkinson. A partial order on $\times 2$-invariant measures. *Mathematical Research Letters*, 15:893–900, 2008.

[37] O. Jenkinson and L. G. Zamboni. Characterisations of balanced words via orderings. *Theoretical Computer Science*, 310:247–271, 2004.

[38] M. Kirkpatrick and M. Slatkin. Superdominance order and distance of trees with bounded maximum degree. *Evolution*, 47(4):1171–1181, 1993.

[39] E. Knuth. *The Art of Computer Programming, Vol 1.* Addison-Wesley, Reading, 1973.

[40] F. T. Leighton. *Introduction to Parallel Algorithms and Architectures: Networks and Algorithms*. Morgan-Kaufmann, 1992.

[41] B. Litow and B. Mans. A note on the Adam conjecture for double loops. *Information Processing Letters*, 66:149–153, 1998.

[42] Z. Lonc, K. Parol, and J. Wojciechowski. On the number of spanning trees in directed circulant graphs. *Networks: An International Journal*, 37:129–133, 2001.

[43] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopaedia of mathematics and its applications*. Cambridge University Press, 2002.

[44] B. Mans. Optimal distributed algorithms in unlabeled tori and chordal rings. *Journal of Parallel and Distributed Computing*, 46(1):80–90, 1997.

[45] M. Morse and G. A. Hedlund. Symbolic dynamics, II: Sturmian trajectories. *American Journal of Mathematics*, 62:1–42, 1940.

[46] R. Neininger. On a multivariate contraction method for random recursive structures with applications to quicksort. *Random Structures and Algorithms*, 19(3–4):498–524, 2001.

[47] R. Neininger. The Wiener index of random trees. *Combinatorics, Probability and Computing*, 11(6):587–597, 2002.

[48] M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Oxford University Press, 1999.

[49] S. Riis. Utilizing public information in network coding (appendix). In Ahlswede, Baumer, Cai, Aydinian, Blinovsky, Deppe, and Mashurian, editors, *General Theory of information Transfer and Combinatorics*, pages 851–882. Springer, Cambridge, 2006.

[50] D. Robinson. Comparison of labeled trees with valency three. *Journal of Combinatorial Theory*, 9:91–116, 1969.

[51] U. Rosler. A limit theorem for 'quicksort'. *Theoretical Informatics and Applications*, 25:85–100, 1991.

[52] E. Schroder. Vier combinatorische probleme. *Zeitschrift fur Mathematik und Physik*, 15:361–376, 1870.

[53] C. Semple and M. Steel. *Phylogenetics*. Oxford University Press, 2003.

[54] B. Slowinski. Probabilities of n-trees under two models: a demonstration that asymmetrical interior nodes are not improbable. *Systematic Zoology*, 39:89–94, 1990.

[55] J. Spies. Some identities involving harmonic numbers. *Mathematics of computation*, 55(192):839–863, 1990.

[56] M. Steel and D. Penny. Distributions of tree comparison metrics-some new results. *Systematic Biology*, 42:126–141, 1993.

[57] L. Székely and H. Wang. On subtrees of trees. *Advances in Applied Mathematics*, 34:138–155, 2005.

[58] Wu T, P. J. Cameron, and S. Riis. On the guessing number of shift graphs. *Journal of discrete algoritms*, 7:220–226, 2009.

[59] J. H. van Lint and R. M. Wilson. *A course in Combinatorics, Second Edition*. Cambridge University Press, 2001.

[60] H. Wang. The extremal values of the wiener index of a tree with given degree sequence. *Discrete Applied Mathematics*, 156:2647–2654, 2008.

[61] H. Wang and T. Wu. The sum of the distances between leaves of a tree and the 'semi-regular' property. *submitted*, 2008.

[62] D. J. A. Welsh. *Complexity: Knots, Colourings and Counting*, volume 186 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 1993.

[63] H Wiener. Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, 69:17–20, 1947.

[64] T. Wu. On regular configurations and disjoint cycles in shift graphs. *Preprint*, 2007.

[65] T. Wu. Cycle packing in shift graphs. *In preparation*, 2008.

[66] G. Yule. A mathematical theory of evolution, based on the conclusion of Dr J.C. Willis. *Philosophical Transactions of the Royal Society of London. Series B*, 213:21–87, 1924.

[67] Y. Zhang, X. Yong, and M. J. Golin. The number of spanning trees in circulant graphs. *Discrete Mathematics*, 223:337–350, 2000.

Acknowledgements

I would like to thank...

...my advisors: Søren Riis and Peter J. Cameron. Søren has been a constant source of support in these years; Peter has taught me a lot about research and academic life. Without their guidance this thesis would not have been possible.

...Andreas Dress, Stefan Grunewald and Peter Erdos for their help and hospitality during my final stage of writing this thesis.

...my committee, Prof. Imre Leader and Prof. Bill Jackson, for their useful comments and encouragement.

...Lihu Xu, Oliver Jenkinson and Jack Douthett for stimulating conversations; Tony Forbes and Harriet Pollats for proof-reading; Carrie Rutherford and Robin Whitty for their support.

...my coauthors: Christoph Buchheim, Peter Cameron, Peter Humphries, Vincent Moulton, Søren Riis, Mike Steel and Hua Wang, for teaching me academic writing and mathematical thinking.

...the people close to the combinatorial study group in Queen Mary and the study group in South Bank. It is a pleasure to discuss mathematics and non-mathematics with them.

...several institutes that I have visited during my PhD study: INI Cambridge, ITCS Tsinghua, PICB Shanghai, Renyi Budapest and MIS Leipzig.

...the teachers who have inspired and encouraged me in the past and all my friends in these years. It is a happy postgraduate journey with the companionship of Mike, Ku, Caifeng, Kurt, Corrado, Yun, Fatma, Emil, Josephine...

...my parents for my existence; my grandmother for her love; uncle Yu Ming for his motivation and support; my brothers and sisters for their understanding; my wife Jean for all of the above and so much more.

Publications

1: *The complexity of the weight problem for permutation groups and matrix groups*
(with P.J. Cameron) Discrete Mathematics, to appear.

2: *On the subgroup distance problem*
(with Ch.Buchheim and P.J. Cameron) Discrete Mathematics, to appear.

3: *On the guessing number of shift graphs*
(with P.J. Cameron and S. Riis) Journal of Discrete Algorithms, to appear.

4: *Refining phylogenetic trees given additional data: An algorithm based on parsimony*
(with V. Moulton and M. Steel) IEEE/ACM Transactions on Computational Biology and Bioinformatics, to appear.

5: *A new characterization of balanced words*
(With P.J. Cameron), WORDS2007, Proceedings of the 6th International Conference on Words, 63–71, 2007.

6: *A topological approach to tree reconstruction*
(with A. Dress and V. Moulton) Submitted.

7: *Barriers in Metric Spaces*
(with A. Dress, V. Moulton and A. Spillner) Submitted.

8: *The sum of the distances between leaves of a tree and the 'semi-regular' property*
(with H. Wang) Submitted.

9: *A note on weight-equivalent graphs*
(with A. Dress and X. Xu) Submitted.